Strumento MiSeqDx™

PER USO DIAGNOSTICO IN VITRO

N. di catalogo DX-410-1001

Uso previsto

Lo strumento di sequenziamento MiSeqDx Illumina misura i segnali fluorescenti di nucleotidi marcati mediante l'uso di reagenti e celle a flusso specifici (kit MiSeqDx Universal 1.0) per lo strumento, hardware di imaging e software per l'analisi dei dati. La piattaforma MiSeqDx è prevista per il sequenziamento mirato del DNA genomico umano ottenuto da campioni di sangue periferico intero. La piattaforma MiSeqDx non è prevista per il sequenziamento dell'intero genoma o il sequenziamento de novo. MiSeqDx è previsto per l'uso con i kit di saggi e reagenti per IVD forniti da Illumina.

Principi della procedura

La piattaforma MiSeqDx Illumina è prevista per il sequenziamento mirato del DNA genomico umano ottenuto da campioni di sangue periferico intero usando i materiali di consumo forniti da Illumina. Il DNA genomico viene elaborato durante la procedura di preparazione della libreria, che specificatamente amplifica le regioni genomiche previste di ciascun campione usando gli oligonucleotidi personalizzati e progettati dall'utente, aggiungendo inoltre gli indici e le sequenze di cattura della cella a flusso ai prodotti amplificati. La preparazione della libreria consiste di quattro fasi principali: ibridazione, estensione-ligazione, amplificazione mediante PCR e normalizzazione della libreria. Le librerie dei campioni così ottenute sono pronte per il sequenziamento sullo strumento MiSeqDx Illumina usando la chimica SBS (sequenziamento mediante sintesi). La chimica SBS utilizza un metodo che fa uso di terminatori reversibili per rilevare le singole basi nucleotidiche man mano che vengono incorporate in filamenti di DNA crescenti. Il software di analisi in tempo reale (RTA) esegue l'analisi delle immagini e l'identificazione delle basi e assegna anche un punteggio qualitativo a ciascuna base per ciascun ciclo di sequenziamento. Al termine dell'analisi primaria, il software MiSeq Reporter sullo strumento MiSeqDx elabora le identificazioni delle basi mediante l'analisi secondaria che include de-multiplexing, generazione di file FASTQ, allineamento, identificazione delle varianti e generazioni di file VCF che contengono le informazioni relative alle varianti individuate in posizioni specifiche in un genoma di riferimento.

Limiti della procedura

1. Per uso diagnostico *in vitro*.
2. Questo prodotto è limitato a fornire:
   - Output di sequenziamento di >1 Gb
   - >3 milioni di letture
   - Lunghezza di lettura (in corse paired-end) 2 x 150 bp
   - >75% delle basi con punteggio qualitativo superiore a Q30 (più del 75% delle basi presentano un punteggio qualitativo su scala Phred superiore a 30, il che indica un'accuratezza di identificazione delle basi superiore al 99,9%)
3. Le varianti in corse per omopolimeri che superano otto basi saranno filtrate nel file VCF (filtro R8).
4. Questo sistema è stato convalidato per il rilevamento di SNV e fino a delezioni di 3 basi. La valutazione delle inserzioni di 1 base è stata limitata a 3 diverse inserzioni su 3 cromosomi separati.
5. Il sistema ha problemi nel rilevare le inserzioni o le delezioni di 1 base in tratti omopolimerici (ad es., PolyA).
6. Il sistema MiSeqDx è progettato per fornire risultati qualitativi (vale a dire genotipo).

7 Come per qualsiasi flusso di lavoro basato sull'ibridazione, polimorfismi o mutazioni latenti nelle regioni che legano gli oligonucleotidi possono interessare gli alleli sottoposti a sonda e, di conseguenza, le identificazioni effettuate.
8 La copertura minima raccomandata per l'amplicone necessaria per l'identificazione accurata delle varianti (Q(max_gt | poly_site) >= 100) è di 75x.

Componenti del prodotto

MiSeqDx Illumina è composto da:

- Strumento MiSeqDx (N. di catalogo DX-410-1001)

I seguenti componenti software sono richiesti per il funzionamento e l'analisi dei dati MiSeqDx:

Applicazione software	Funzione	Descrizione
MOS - software operativo MiSeqDx	Controlla il funzionamento dello strumento	L'applicazione software MOS gestisce il funzionamento dello strumento durante il sequenziamento e genera immagini da usare con il software di analisi in tempo reale (RTA).
RTA - software di analisi in tempo reale	Esegue l'analisi primaria	L'applicazione software RTA converte le immagini generate da MOS per ciascuna tile per ciclo della corsa di sequenziamento in file di identificazioni delle basi che rappresentano gli input per il software MiSeq Reporter. L'applicazione software RTA non contiene un'interfaccia utente.
MiSeq Reporter	Esegue l'analisi secondaria	Il software MiSeq Reporter elabora le identificazioni delle basi mediante l'analisi secondaria che include de-multiplexing, generazione di file FASTQ, allineamento, identificazione delle varianti e generazioni di file VCF che contengono le informazioni relative alle varianti individuate in posizioni specifiche in un genoma di riferimento.

Conservazione e manipolazione

Elemento	Specifica
Temperatura	Trasporto e conservazione: da -10 °C a 40 °C Condizioni di funzionamento: da 19 °C a 25 °C
Umidità	Trasporto e conservazione: umidità senza condensa Condizioni di funzionamento: umidità relativa 30-75% (senza condensa)

Apparecchiatura e materiali richiesti, non forniti

Materiali di consumo per il sequenziamento

Kit MiSeqDx Universal 1.0 (n. di catalogo DX-103-1001)

Materiali di consumo forniti dall'utente

Prima di avviare una corsa di sequenziamento, accertarsi di avere a disposizione i seguenti materiali di consumo forniti dall'utente.

Materiale di consumo	Scopo
Salviettine imbevute di alcol isopropilico al 70% oppure di etanolo al 70%	Pulizia del vano portacella
Panno da laboratorio a bassissimo rilascio di particelle	Pulizia del piano portacelle

Materiale di consumo	Scopo
Carta pulente per lenti, 10x15 cm ca.	Pulizia della cella a flusso
Tween 20	Lavaggio dello strumento
Pinzette di plastica con punta quadrata (opzionale)	Estrazione della cella a flusso dal relativo contenitore
Acqua, priva di DNasi, priva di RNasi	Lavaggio dello strumento

Linee guida per l'acqua priva di DNasi e RNasi

Per eseguire le procedure dello strumento usare sempre acqua priva di DNasi e RNasi. Non usare mai acqua di rubinetto o acqua deionizzata. A titolo di esempio, è accettabile quanto segue:

- Acqua pari a 18 MΩ (Megaohm)
- Acqua Milli-Q
- Acqua Super-Q
- Acqua sterile per biologia molecolare

Avvertenze e precauzioni

ATTENZIONE
La legge federale limita la vendita di questo dispositivo da parte o dietro prescrizione di un medico o di un medico autorizzato dalla legge dello stato in cui esercita, ad usare o ad ordinare l'uso del dispositivo.

1. Alcuni componenti di reagenti forniti da Illumina da usare con lo strumento MiSeqDx contengono formammide, una ammide alifatica che è una probabile tossina riproduttiva. Per maggiori informazioni, vedere gli inserti della confezione dei rispettivi prodotti. L'inalazione, l'ingestione, il contatto con la pelle o con gli occhi possono causare lesioni personali. Smaltire i contenitori e i contenuti non utilizzati conformemente alle norme di sicurezza in vigore localmente. Per informazioni, contattare l'Assistenza tecnica Illumina.
2. Alcuni componenti dei reagenti forniti da Illumina contengono 2-mercaptoetanolo, un agente riducente. Per maggiori informazioni, vedere gli inserti della confezione dei rispettivi prodotti. L'inalazione, l'ingestione, il contatto con la pelle o con gli occhi possono causare lesioni personali. Utilizzarlo in un'area ben ventilata e smaltire i contenitori e i contenuti non utilizzati conformemente alle norme di sicurezza in vigore localmente. Per informazioni, contattare l'Assistenza tecnica Illumina.
3. Maneggiare tutti campioni come se fossero agenti potenzialmente infettivi.
4. Il mancato rispetto delle procedure descritte può produrre risultati errati o una riduzione significativa della qualità del campione.
5. Adottare le normali precauzioni di laboratorio. Non pipettare con la bocca. Non mangiare, bere o fumare nelle zone designate per il lavoro. Manipolare i campioni e i reagenti del kit indossando guanti e indumenti da laboratorio monouso. Dopo aver maneggiato i campioni e i reagenti del kit lavarsi bene le mani.
6. È necessario adottare idonee pratiche di laboratorio e una valida igiene per impedire la contaminazione di reagenti, strumenti e campioni di DNA genomico con i prodotti della PCR. La contaminazione da PCR può produrre risultati inesatti e inaffidabili.
7. Al fine di prevenire la contaminazione, accertarsi che le aree di pre-amplificazione e post-amplificazione siano dotate delle apposite attrezzature (ad es., pipette, punte di pipette, agitatori e centrifughe).
8. Le combinazioni indice-campione devono corrispondere esattamente a quelle del foglio campioni. La non corrispondenza fra il foglio campioni e il layout della piastra risulterà in una perdita di identificazione dei campioni positivi e nella refertazione di risultati errati.
9. Durante la fase di normalizzazione della libreria in base all'inserto della confezione dei rispettivi reagenti, è estremamente importante risospendere completamente il pellet di microsfere della libreria. Questo è fondamentale per ottenere una densità dei cluster omogenea sulla cella a flusso MiSeqDx.

10 Attenersi ai tempi di incubazione indicati nella procedura di normalizzazione della libreria come descritto nell'inserto della confezione dei rispettivi reagenti. Un'incubazione inadeguata può influire sulla rappresentazione della libreria e sulla densità dei cluster.
11 Si consiglia vivamente l'installazione di un software antivirus, fornito dall'utente, per proteggere il computer dai virus. Consultare il Manuale d'uso per le istruzioni relative all'installazione.
12 Non far funzionare MiSeqDx se un qualsiasi pannello è rimosso. Il funzionamento dello strumento con un qualsiasi pannello rimosso crea esposizione potenziale a tensioni di linea e tensioni c.c..
13 Non toccare il piano portacelle nello scomparto della cella a flusso. Il riscaldatore in questo scomparto funziona a una temperatura compresa tra 22 °C e 95 °C e potrebbe provocare bruciature.
14 Lo strumento pesa circa 58 kg ed eventuali cadute o errata manipolazione possono causare lesioni gravi.

Note sulle procedure

Il rendimento per una corsa MiSeqDx può essere tra 8 e 48 campioni. I primer di indicizzazione usati durante l'amplificazione mediante PCR devono essere scelti in base al rendimento dei campioni finale desiderato per garantire diversità nella sequenza d'indice.

NOTA
Per la massima efficienza di rendimento, eseguire la preparazione della libreria per un massimo di 96 campioni e quindi dividere i campioni in due corse di sequenziamento con un massimo di 48 campioni ciascuno.

MiSeqDx utilizza un LED verde per il sequenziamento delle basi G/T e un LED rosso per il sequenziamento delle basi A/C. Per garantire la corretta registrazione, a ogni corsa deve essere letto almeno uno dei due nucleotidi per ogni canale cromatico. È importante conservare l'equilibrio cromatico per ogni base dell'indice letta sottoposta a sequenziamento, altrimenti potrebbe verificarsi un problema di registrazione durante il sequenziamento della Lettura indici.

Se si sequenziano meno di 48 campioni in una corsa di sequenziamento, selezionare gli indici appropriati in base alle relative sequenze per conservare il bilanciamento cromatico nei canali verde e rosso. Come requisito minimo, le corse da 8 a 48 campioni devono includere una combinazione di primer di indicizzazione identificata nell'inserto della confezione del kit MiSeqDx Universal 1.0.

Per elaborare in maniera accurata corse più piccole, devono essere presenti almeno otto campioni. Se non sono disponibili sei campioni univoci (esclusi i controlli positivi e negativi), è accettabile completare la corsa con replicati o qualsiasi campione di DNA genomico umano. Per il set minimo di indici bilanciati per colore da usare nelle corse di sequenziamento a otto campioni, vedere l'inserto della confezione del kit MiSeqDx Universal 1.0.

Istruzioni per l'uso

Le istruzioni seguenti per l'uso dello strumento MiSeqDx richiedono i reagenti forniti nel kit MiSeqDx Universal 1.0.

Preparazione del foglio campioni per MiSeqDx

1 Nella schermata Welcome (Benvenuto) di Worklist Manager Illumina, selezionare **Create Worklist** (Crea lista di lavoro).
2 Nel campo Test Type (Tipo di test), selezionare **MiSeqDx Universal**.
3 Nel campo Worklist Name (Nome lista di lavoro), immettere il nome per il foglio campioni.
   — Se per il nome del foglio campioni viene utilizzato l'ID alfanumerico del codice a barre della cartuccia di reagenti, il software operativo MiSeq (MOS) troverà automaticamente il foglio campioni.
   — Se per il foglio campioni viene utilizzato un qualsiasi altro nome, il pulsante **Browse** (Sfoglia) nel software operativo MiSeq (MOS) può essere utilizzato per trovare il foglio campioni corretto.
4 [Opzionale] Immettere una descrizione per identificare la corsa.
5 Assicurarsi che la data corrisponda alla data di inizio della corsa.
6 Selezionare **Next** (Avanti).

Inserimento di informazioni sui campioni

1. Nella scheda Table (Tabella) o nella scheda Plate (Piastra), inserire le seguenti informazioni per ogni pozzetto contenente campione:
   a. **Sample ID** (ID campione): inserire un ID campione univoco.
   b. **Index 1 e Index 2** (Indice 1 e Indice 2): specificare l'adattatore dell'indice che verrà utilizzato per ogni Lettura indici.
   c. **Manifest** (Manifest): specificare in nome del file manifest che contiene le informazioni relative ai campioni in quel determinato pozzetto.
2. [Opzionale] Per registrare informazioni più dettagliate sui campioni, inserire il nome e la descrizione di un campione.
3. [Opzionale] Per identificare i controlli sulla piastra, selezionare Negative (Negativo) o Positive (Positivo) dal menu a discesa **Control** (Controllo).
4. Andare alla scheda Plate Graphic (Schema piastra) e utilizzare l'opzione **Copy to Clipboard** (Copia in appunti) o **Print** (Stampa) per acquisire un'immagine della piastra campioni.
5. Selezionare **Fine**.

Preparazione dei campioni

La procedura seguente deve essere eseguita in base alle Istruzioni per l'uso contenute nell'inserto della confezione del kit MiSeqDx Universal 1.0:

- Ibridazione del pool di oligonucleotidi
- Rimozione degli oligonucleotidi non legati
- Estensione-ligazione degli oligonucleotidi legati
- Amplificazione mediante PCR
- Pulizia della PCR
- Normalizzazione della libreria
- Creazione di pool di librerie

Preparazione della cartuccia di reagenti

1. Scongelare la cartuccia di reagenti MiSeqDx in un bagnomaria contenente abbastanza acqua deionizzata a temperatura ambiente così da immergere la base della cartuccia di reagenti fino alla linea di livello acqua stampata sulla cartuccia stessa. Evitare che l'acqua superi la linea di massimo livello acqua.
2. Lasciare la cartuccia di reagenti a scongelare a bagnomaria a temperatura ambiente per circa un'ora o fino a completo scongelamento.
3. Rimuovere la cartuccia dal bagno d'acqua e batterla delicatamente sul banco per far fuoriuscire l'acqua in eccesso dalla base. Asciugare la base della cartuccia. Verificare che sulla parte superiore della cartuccia di reagenti non sia caduta dell'acqua.

Ispezione della cartuccia di reagenti

1. Capovolgere la cartuccia dieci volte per miscelare i reagenti scongelati e poi ispezionare visivamente tutti i serbatoi per accertarsi che siano scongelati.

NOTA
È fondamentale che i reagenti nella cartuccia siano scongelati completamente e miscelati per assicurare il sequenziamento corretto.

2. Ispezionare visivamente il reagente nella posizione 1 per accertarsi che sia ben miscelato e privo di precipitati.
3. Battere delicatamente la cartuccia sul banco per ridurre le bolle d'aria nei reagenti.

NOTA
I tubi dei pescanti del MiSeqDx vanno fino al fondo di ciascun serbatoio per aspirare i reagenti, per questa ragione è importante che i serbatoi non contengano bolle.

4 Riporre la cartuccia in ghiaccio o conservarla a una temperatura compresa tra 2 °C e 8 °C (fino a 6 ore) finché non si è pronti a impostare la corsa. Per risultati ottimali, procedere direttamente caricando il campione e impostando la corsa.

Preparazione dei campioni per il sequenziamento

1. Portare il tampone di diluizione libreria a temperatura ambiente. Agitare il tampone di diluizione libreria energicamente e assicurarsi che tutti i precipitati si siano completamente sciolti.
2. Se la piastra **SGP** è stata conservata congelata, scongelarla fino a raggiungere la temperatura ambiente.
3. Centrifugare la piastra **SGP** a 1.000 x g a 20 °C per 1 minuto.
4. Predisporre una provetta Eppendorf nuova (d'ora in avanti indicata come provetta **PAL**).
5. Determinare i campioni da raggruppare in pool per il sequenziamento. Per il sequenziamento è possibile utilizzare al massimo 48 campioni contemporaneamente.
6. Se la piastra **SGP** è stata conservata congelata, mescolare ciascuna libreria da sequenziare pipettando su e giù per 3-5 volte.
7. Trasferire 5 µl di ciascuna libreria da sequenziare dalla piastra **SGP**, colonna per colonna, a una striscia a otto provette per PCR. Sigillare l'**SGP** con un sigillo adesivo per piastre e metterla da parte.

NOTA
Dopo l'uso, conservare la piastra **SGP** fra -25 °C e -15 °C. La piastra **SGP** sigillata rimane stabile per un massimo di 3 giorni.

8. Combinare e trasferire il contenuto della striscia a otto provette per PCR nella provetta **PAL**. Miscelare energicamente la provetta **PAL**.
9. Predisporre una provetta Eppendorf nuova (d'ora in avanti indicata come provetta **DAL**).
10. Dispensare 585 µl di tampone di diluizione libreria nella provetta **DAL**.
11. Dispensare 6 µl di controllo interno PhiX 20 pM nella provetta **DAL**. Pipettare su e giù per 3-5 volte per sciacquare la punta e assicurare che il trasferimento sia completo.
12. Trasferire 9 µl di **PAL** nella provetta **DAL** contenente il tampone di diluizione libreria. Pipettare su e giù per 3-5 volte per sciacquare la punta e assicurare che il trasferimento sia completo.
13. Mescolare la **DAL** agitando la provetta molto velocemente.
14. Centrifugare la provetta **DAL** a 1.000 x g a 20 °C per 1 minuto.
15. Incubare la provetta **DAL** in un blocco termico a 96 °C per 2 minuti.
16. Dopo l'incubazione, capovolgere la provetta **DAL** 1-2 volte per mescolarla, quindi porla immediatamente in un bagno di acqua e ghiaccio.
17. Tenere la provetta **DAL** nel bagno di acqua e ghiaccio per 5 minuti.

Caricamento delle librerie di campioni nella cartuccia

1. Utilizzare la punta di una pipetta pulita e vuota da 1 ml per forare la capsula di alluminio che sigilla il serbatoio contrassegnato con la dicitura **Load Samples** (Caricamento campioni).
2. Pipettare 600 µl delle librerie di campioni nel serbatoio contrassegnato con la dicitura **Load Samples** (Caricamento campioni). Fare attenzione a evitare di toccare la capsula sigillante durante l'erogazione del campione.
   Una volta caricato il campione, verificare la presenza di bolle d'aria nel serbatoio. In caso di presenza di bolle d'aria, picchiettare delicatamente la cartuccia sul banco in modo da farle fuoriuscire.
3. Passare direttamente alla procedura di impostazione della corsa usando l'interfaccia del software operativo MiSeq (MOS).

Impostazione della corsa

1. Eseguire l'accesso al software operativo MiSeq (MOS)..
2. Selezionare **Sequence** (Sequenziamento).
   Si aprono una serie di schermate di configurazione nell'ordine seguente: Select Run Type, Load Flow Cell, Load Reagents, Review e Pre-Run Check (Selezione tipo di corsa, Caricamento cella a flusso, Caricamento reagenti, Revisione e Verifica pre-corsa).
3. Selezionare **Diagnostic Run** (Corsa diagnostica).
4. Quando viene visualizzata la schermata Load Flow Cell (Carica cella a flusso), pulire e quindi caricare la cella a flusso.
5. Chiudere il coperchio a scatto e lo sportello dello scomparto della cella a flusso.
   Il coperchio a scatto e lo sportello dello scomparto della cella a flusso devono essere chiusi per poter avviare la corsa. Una volta eseguito il caricamento della cella a flusso, il software legge l'etichetta RFID e ne registra il contenuto. Se l'etichetta RFID viene letta correttamente, nell'angolo inferiore destro della schermata appare un messaggio di conferma.
6. Quando viene visualizzata la schermata Load Reagents (Caricamento reagenti), svuotare il flacone degli scarti, caricare un flacone di soluzione SBS MiSeqDx (PR2) e quindi caricare la cartuccia di reagenti.
   Quando il flacone di Soluzione SBS MiSeqDx (PR2) e la cartuccia di reagenti sono caricate, il software legge e registra le etichette RFID. Se l'etichetta RFID viene letta correttamente, nell'angolo inferiore destro della schermata appare un messaggio di conferma.
7. Selezionare il foglio campioni appropriato.
   Per impostazione predefinita, il software cercherà il file del foglio campioni con un nome che corrisponde al numero del codice a barre della cartuccia di reagenti caricata sullo strumento.
8. Confermare le impostazioni della corsa e i risultati della verifica pre-corsa.
9. Avviare la corsa.
   La schermata Sequencing (Sequenziamento) si apre all'inizio della corsa. La schermata fornisce una rappresentazione visiva dell'avanzamento della corsa, che comprende le intensità e i punteggi qualitativi (Q-scores).

Risultati

Il software integrato di analisi primaria in tempo reale (RTA) esegue l'analisi delle immagini e l'identificazione delle basi e assegna anche un punteggio qualitativo a ciascuna base per ciascun ciclo di sequenziamento. Al termine dell'analisi primaria, il software MiSeq Reporter sullo strumento MiSeqDx inizia l'analisi secondaria, come descritto di seguito.

De-multiplexing

Il de-multiplexing costituisce la prima fase dell'analisi se nel foglio campioni sono elencati più campioni e la corsa presenta letture indici. Mediante il de-multiplexing è possibile separare i dati da campioni raggruppati in pool in base a sequenze indice brevi che etichettano i campioni di diverse librerie. Ciascuna sequenza di lettura indici e viene messa a confronto con le sequenze indice specificate nel foglio campioni. In questa fase non vengono considerati i valori qualitativi.

Generazione di file FASTQ

Dopo il de-multiplexing, MiSeq Reporter genera file intermedi in formato FASTQ, un formato di testo utilizzato per rappresentare le sequenze. I file FASTQ contengono le letture di ciascun campione e i punteggi qualitativi, con l'esclusione delle letture dei cluster che non hanno attraversato il filtro. Il punteggio qualitativo Q è calcolato come $-10 \log_{10} P$, dove P è la probabilità di identificazione delle basi errata.

Allineamento

Mediante l'allineamento è possibile confrontare le sequenze rispetto a un riferimento al fine di identificare una relazione fra le sequenze e assegnare un punteggio in base a regioni di similarità. Le letture allineate vengono scritte su file in formato BAM. Per il kit MiSeqDx Universal 1.0, MiSeq Reporter utilizza un algoritmo di Smith-Waterman con matrice a banda che esegue allineamenti locali di sequenze per determinare il grado di similarità fra due sequenze.

Identificazione delle varianti

Per il kit MiSeqDx Universal 1.0, MiSeq Reporter utilizza il Variant Caller Starling, che identifica SNP e Indel piccoli, e riassume la profondità e le probabilità per ciascun sito nel genoma. Starling produce report di SNP e Indel in formato html e file di testo delimitato da tabulazione contenenti le varianti in formato VCF (Variant Call Format). Per informazioni su come possono essere calcolati i risultati dai file VCF, vedere la *Guida per l'utente di MiSeq Reporter (n. codice 15039188)*.

Caratteristiche prestazionali

Accuratezza

Sono stati condotti tre studi separati per valutare l'accuratezza della piattaforma MiSeqDx.

Studio 1

Questo studio usa un saggio rappresentativo progettato per interrogare una varietà di geni che coprono 24.434 basi su 19 diversi cromosomi e che contengono potenzialmente gli esoni rilevanti dal punto di vista clinico. I 13 campioni unici usati in questo studio appartengono a due genitori e 11 bambini. Questi campioni sono spesso stati sottoposti a sequenziamento da laboratori multipli e metodi di sequenziamento multipli. Sei campioni sono femminili e sette sono maschili. L'accuratezza è stata determinata per le varianti di singolo nucleotide (SNV) confrontando i dati dello studio su un database dei riferimenti ben caratterizzati. Le sequenze di riferimento del database erano derivate dalla combinazione di metodologie di sequenziamento multiple, dati disponibili pubblicamente e informazioni ereditarie. Per valutare l'accuratezza del sistema la tabella seguente è stata compilata in base ai dati ottenuti dalla prima corsa nello studio. Per questo studio non sono stati ripetuti test.

I risultati ottenuti da questo studio sono presentati qui di seguito.

Tabella 1 Studio 1 - Accuratezza dei dati a livello di ampliconi per la piattaforma MiSeqDx

Amplicone	Cromosoma	Dimensione del frammento analizzato[1]	Contenuto genomico dell'amplicone	N. di campioni unici	N. totale di campioni analizzati[2]	N. di identificazioni per campione che è stato possibile effettuare[3]	N. di identificazioni non riuscite[4]	N. di identificazioni corrette per campione[5]	N. di identificazioni errate[6]	% di identificazioni corrette[7]
1	1	132	Poly C (5), 63% GC	13	15	132	0	132	0	100
2	1	128	Poly T (5)	13	15	128	0	128	0	100
3	2	133	-	13	15	133	0	133	0	100
4	2	119	-	13	15	119	0	119	0	100
5	2	127	Poly T (5)	13	15	127	0	127	0	100
6	2	135	Poly A (6)	13	15	135	0	135	0	100
7	2	122	Poly T (5), Poly C (5)	13	15	122	0	122	0	100
8	2	110	Poly T (5)	13	15	110	0	110	0	100
9[8]	2	131	Poly A (14)	13	15	130-131	0	130-131	9	99,54
10	2	117	-	13	15	117	0	117	0	100
11	2	121	-	13	15	121	0	121	0	100
12	2	114	-	13	15	114	0	114	0	100
13	2	129	Poly A (5)	13	15	129	0	129	0	100
14	3	131	Poly A (5), Poly T (5)	13	15	131	0	131	0	100
15	3	130	-	13	15	130	0	130	0	100

Ampli-cone	Cromo-soma	Dimensione del frammento analizzato[1]	Contenuto genomico dell'amplicone	N. di cam-pioni unici	N. totale di cam-pioni analizzati[2]	N. di identificazioni per campione che è stato possibile effettuare[3]	N. di identifi-cazioni non riuscite[4]	N. di identifi-cazioni corrette per campione[5]	N. di identifi-cazioni errate[6]	% di identificazioni corrette[7]
16	3	130	-	13	15	130	0	130	0	100
17	3	117	-	13	15	117	0	117	0	100
18	3	136	Poly T (5)	13	15	136	0	136	0	100
19	3	131	Poly T (5), SNV	13	15	131	0	131	0	100
20	3	123	Poly A (5)	13	15	123	0	123	0	100
21	3	117	Poly A (6), Poly T (5), Regioni omologhe su un cromosoma diverso	13	15	117	0	117	0	100
22	3	119	Regioni omologhe su un cromosoma diverso	13	15	119	0	119	0	100
23	3	120	-	13	15	120	0	120	0	100
24	3	129	Poly T (5)	13	15	129	0	129	0	100
25	4	133	Poly C (7), 66% GC	13	15	133	0	133	0	100
26	4	135	Poly C (5), 69% GC	13	15	135	0	135	0	100

Ampli-cone	Cromo-soma	Dimensione del frammento analizzato[1]	Contenuto genomico dell'amplicone	N. di cam-pioni unici	N. totale di cam-pioni analizzati[2]	N. di identificazioni per campione che è stato possibile effettuare[3]	N. di identifi-cazioni non riuscite[4]	N. di identifi-cazioni corrette per campione[5]	N. di identifi-cazioni errate[6]	% di identificazioni corrette[7]
27	4	123	SNV	13	15	123	0	123	0	100
28	4	134	-	13	15	134	0	134	0	100
29	4	132	-	13	15	132	0	132	0	100
30	4	121	Poly A (5), SNV	13	15	121	0	121	0	100
31	4	125	-	13	15	125	0	125	0	100
32	4	134	Poly T (5)	13	15	134	0	134	0	100
33	4	118	-	13	15	118	0	118	0	100
34	4	122	Poly A (5)	13	15	122	0	122	0	100
35	4	131	-	13	15	131	0	131	0	100
36	4	133	-	13	15	133	0	133	0	100
37	4	128	Poly T (6)	13	15	128	0	128	0	100
38	4	131	-	13	15	131	0	131	0	100
39	4	129	Poly A (5), Poly T (5), SNV	13	15	129	0	129	0	100
40	4	133	Poly T (5), SNV	13	15	133	0	133	0	100
41	4	112	SNV	13	15	112	0	112	0	100

Ampli-cone	Cromo-soma	Dimensione del frammento analizzato[1]	Contenuto genomico dell'amplicone	N. di cam-pioni unici	N. totale di cam-pioni analizzati[2]	N. di identificazioni per campione che è stato possibile effettuare[3]	N. di identifi-cazioni non riuscite[4]	N. di identifi-cazioni corrette per campione[5]	N. di identifi-cazioni errate[6]	% di identificazioni corrette[7]
42	4	133	-	13	15	133	0	133	0	100
43	4	135	-	13	15	135	0	135	0	100
44	4	122	-	13	15	122	0	122	0	100
45	4	117	-	13	15	117	0	117	0	100
46[9]	4	124	-	13	15	125	0	125	0	100
47	4	117	Poly T (5)	13	15	117	0	117	0	100
48	4	128	Poly A (7)	13	15	128	0	128	0	100
49	4	123	Poly A (6)	13	15	123	0	123	0	100
50	4	133	-	13	15	133	0	133	0	100
51	4	112	-	13	15	112	0	112	0	100
52	4	129	-	13	15	129	0	129	0	100
53	4	126	-	13	15	126	0	126	0	100
54	4	132	-	13	15	132	0	132	0	100
55	5	131	-	13	15	131	0	131	0	100
56	5	119	-	13	15	119	0	119	0	100
57	5	120	Poly A (5)	13	15	120	0	120	0	100
58	5	119	-	13	15	119	0	119	0	100
59	5	118	-	13	15	118	0	118	0	100

Ampli-cone	Cromo-soma	Dimensione del frammento analizzato[1]	Contenuto genomico dell'amplicone	N. di cam-pioni unici	N. totale di cam-pioni analizzati[2]	N. di identificazioni per campione che è stato possibile effettuare[3]	N. di identifi-cazioni non riuscite[4]	N. di identifi-cazioni corrette per campione[5]	N. di identifi-cazioni errate[6]	% di identificazioni corrette[7]
60	5	112	-	13	15	112	0	112	0	100
61	5	120	-	13	15	120	0	120	0	100
62	5	120	Poly A (5)	13	15	120	0	120	0	100
63	5	115	CT(5)	13	15	115	0	115	0	100
64	5	112	SNV	13	15	112	0	112	0	100
65	5	135	Poly T (6)	13	15	135	0	135	0	100
66	5	131	63% GC	13	15	131	0	131	0	100
67	5	121	-	13	15	121	0	121	0	100
68	5	132	Poly A (6), Poly T (8)	13	15	132	0	132	0	100
69	7	133	-	13	15	133	0	133	0	100
70	7	120	60% GC	13	15	120	0	120	0	100
71	7	135	-	13	15	135	0	135	0	100
72	7	126	Poly A (5), 59% GC	13	15	126	0	126	0	100
73	7	134	-	13	15	134	0	134	0	100
74	7	122	Poly C (5), 63% GC	13	15	122	0	122	0	100

Amplicone	Cromosoma	Dimensione del frammento analizzato[1]	Contenuto genomico dell'amplicone	N. di campioni unici	N. totale di campioni analizzati[2]	N. di identificazioni per campione che è stato possibile effettuare[3]	N. di identificazioni non riuscite[4]	N. di identificazioni corrette per campione[5]	N. di identificazioni errate[6]	% di identificazioni corrette[7]
75	7	127	59% GC; SNV	13	15	127	0	127	0	100
76	7	123	-	13	15	123	0	123	0	100
77	7	125	-	13	15	125	0	125	0	100
78	7	133	Poly A (5), Poly T (5)	13	15	133	0	133	0	100
79	7	116	-	13	15	116	0	116	0	100
80	7	135	-	13	15	135	0	135	0	100
81	7	118	-	13	15	118	0	118	0	100
82	7	136	67% GC	13	15	136	0	136	0	100
83	7	131	58% GC	13	15	131	0	131	0	100
84	7	119	Poly G (6), 61% GC	13	15	119	0	119	0	100
85	7	122	Poly T (5)	13	15	122	0	122	0	100
86	7	123	Poly A (6)	13	15	123	0	123	0	100
87	8	127	60% GC	13	15	127	0	127	0	100
88	8	129	57% GC	13	15	129	0	129	0	100
89	9	130	Poly T (5)	13	15	130	0	130	0	100
90	9	116	-	13	15	116	0	116	0	100

Ampli-cone	Cromo-soma	Dimensione del frammento analizzato[1]	Contenuto genomico dell'amplicone	N. di cam-pioni unici	N. totale di cam-pioni analizzati[2]	N. di identificazioni per campione che è stato possibile effettuare[3]	N. di identifi-cazioni non riuscite[4]	N. di identifi-cazioni corrette per campione[5]	N. di identifi-cazioni errate[6]	% di identificazioni corrette[7]
91	9	119	Regioni omologhe su un cromosoma diverso	13	15	119	0	119	0	100
92	9	121	-	13	15	121	0	121	0	100
93	9	117	Regioni omologhe su un cromosoma diverso	13	15	117	0	117	0	100
94	9	114	-	13	15	114	0	114	0	100
95[10]	9	129	Poly A (14)	13	15	130	0	129 (di 130)	15	99,23
96	9	114	Regioni omologhe su un cromosoma diverso; SNV	13	15	114	0	114	0	100
97	9	122	-	13	15	122	0	122	0	100
98	9	127	Poly A (5), Poly C (5)	13	15	127	0	127	0	100
99	9	133	-	13	15	133	0	133	0	100
100	9	138	64% GC	13	15	138	0	138	0	100
101	9	139	-	13	15	139	0	139	0	100
102	9	116	-	13	15	116	0	116	0	100

Ampli-cone	Cromo-soma	Dimensione del frammento analizzato[1]	Contenuto genomico dell'amplicone	N. di cam-pioni unici	N. totale di cam-pioni analizzati[2]	N. di identificazioni per campione che è stato possibile effettuare[3]	N. di identifi-cazioni non riuscite[4]	N. di identifi-cazioni corrette per campione[5]	N. di identifi-cazioni errate[6]	% di identificazioni corrette[7]
103	9	133	Poly A (5), 57% GC	13	15	133	0	133	0	100
104	9	138	57% GC	13	15	138	0	138	0	100
105	9	136	Poly C (5), 67% GC	13	15	136	0	136	0	100
106	9	118	70% GC	13	15	118	0	118	0	100
107	10	128	62% GC	13	15	128	0	128	0	100
108	10	120	60% GC	13	15	120	0	120	0	100
109	10	139	58% GC; SNV	13	15	139	0	139	0	100
110	10	118	57% GC	13	15	118	0	118	0	100
111	10	123	Poly T (5)	13	15	123	0	123	0	100
112	10	121	-	13	15	121	0	121	0	100
113	10	129	26% GC	13	15	129	0	129	0	100
114	10	122	-	13	15	122	0	122	0	100
115	10	124	Poly T (5); Regioni omologhe su un cromosoma diverso	13	15	124	0	124	0	100
116	10	135	CA(4)	13	15	135	0	135	0	100

Amplicone	Cromosoma	Dimensione del frammento analizzato[1]	Contenuto genomico dell'amplicone	N. di campioni unici	N. totale di campioni analizzati[2]	N. di identificazioni per campione che è stato possibile effettuare[3]	N. di identificazioni non riuscite[4]	N. di identificazioni corrette per campione[5]	N. di identificazioni errate[6]	% di identificazioni corrette[7]
117	10	135	Poly A (6); Regioni omologhe su un cromosoma diverso	13	15	135	0	135	0	100
118	10	119	Poly C (5); SNV	13	15	119	0	119	0	100
119	10	125	-	13	15	125	0	125	0	100
120	10	131	-	13	15	131	0	131	0	100
121	10	117	-	13	15	117	0	117	0	100
122	10	116	-	13	15	116	0	116	0	100
123	10	129	58% GC	13	15	129	0	129	0	100
124	11	117	Poly T (10)	13	15	117	0	117	0	100
125	11	117	Poly T (5)	13	15	117	0	117	0	100
126	11	113	Poly A (5)	13	15	113	0	113	0	100
127	11	129	-	13	15	129	0	129	0	100
128	11	121	Poly T (5)	13	15	121	0	121	0	100
129	11	123	-	13	15	123	0	123	0	100
130	11	127	Poly A (6)	13	15	127	0	127	0	100
131	11	136	Poly T (6)	13	15	136	0	136	0	100

Amplicone	Cromosoma	Dimensione del frammento analizzato[1]	Contenuto genomico dell'amplicone	N. di campioni unici	N. totale di campioni analizzati[2]	N. di identificazioni per campione che è stato possibile effettuare[3]	N. di identificazioni non riuscite[4]	N. di identificazioni corrette per campione[5]	N. di identificazioni errate[6]	% di identificazioni corrette[7]
132	11	132	Poly T (5)	13	15	132	0	132	0	100
133	11	115	-	13	15	115	0	115	0	100
134	11	117	Poly T (8); 19% GC	13	15	117	0	117	0	100
135	11	134	Poly A (5); Poly T (5)	13	15	134	0	134	0	100
136	11	131	Poly A (5)	13	15	131	0	131	0	100
137	11	133	26% GC; SNV	13	15	133	0	133	0	100
138	11	137	Poly T (8); SNV	13	15	137	0	137	0	100
139	11	131	Poly A (5)	13	15	131	0	131	0	100
140	12	131	-	13	15	131	0	131	0	100
141	12	128	-	13	15	128	0	128	0	100
142	12	133	Poly A (5)	13	15	133	0	133	0	100
143	12	136	-	13	15	136	0	136	0	100
144	12	124	-	13	15	124	0	124	0	100
145	12	122	59% GC	13	15	122	0	122	0	100
146	13	122	-	13	15	122	0	122	0	100

Ampli-cone	Cromo-soma	Dimensione del frammento analizzato[1]	Contenuto genomico dell'amplicone	N. di cam-pioni unici	N. totale di cam-pioni analizzati[2]	N. di identificazioni per campione che è stato possibile effettuare[3]	N. di identifi-cazioni non riuscite[4]	N. di identifi-cazioni corrette per campione[5]	N. di identifi-cazioni errate[6]	% di identificazioni corrette[7]
147	13	116	Poly C (5)	13	15	116	0	116	0	100
148	13	133	-	13	15	133	0	133	0	100
149	13	117	SNV	13	15	117	0	117	0	100
150	13	124	Poly T (6)	13	15	124	0	124	0	100
151	13	123	Poly T (5); 26% GC	13	15	123	0	123	0	100
152	13	115	Poly A (5)	13	15	115	0	115	0	100
153	13	125	-	13	15	125	0	125	0	100
154	13	121	-	13	15	121	0	121	0	100
155	13	123	-	13	15	123	0	123	0	100
156	13	114	-	13	15	114	0	114	0	100
157	13	119	-	13	15	119	0	119	0	100
158	14	122	58% GC	13	15	122	0	122	0	100
159	16	122	-	13	15	122	0	122	0	100
160	16	121	-	13	15	121	0	121	0	100
161	16	123	Poly C (5)	13	15	123	0	123	0	100
162	17	119	-	13	15	119	0	119	0	100
163	17	119	61% GC	13	15	119	0	119	0	100

Ampli-cone	Cromo-soma	Dimensione del frammento analizzato[1]	Contenuto genomico dell'amplicone	N. di cam-pioni unici	N. totale di cam-pioni analizzati[2]	N. di identificazioni per campione che è stato possibile effettuare[3]	N. di identifi-cazioni non riuscite[4]	N. di identifi-cazioni corrette per campione[5]	N. di identifi-cazioni errate[6]	% di identificazioni corrette[7]
164	17	135	-	13	15	135	0	135	0	100
165	17	116	Poly C (6); 60% GC; SNV	13	15	116	0	116	0	100
166	17	123	-	13	15	123	0	123	0	100
167	17	116	62% GC	13	15	116	0	116	0	100
168	17	118	Poly C (5); 65% GC	13	15	118	0	118	0	100
169	17	129	-	13	15	129	0	129	0	100
170	17	131	Poly G (6); 67% GC; SNV	13	15	131	0	131	0	100
171	17	127	61% GC	13	15	127	0	127	0	100
172	17	118	Poly C (5)	13	15	118	0	118	0	100
173	17	138	61% GC	13	15	138	0	138	0	100
174	17	131	58% GC	13	15	131	0	131	0	100
175	18	112	-	13	15	112	0	112	0	100
176	18	124	-	13	15	124	0	124	0	100
177	18	134	Poly A (6)	13	15	134	0	134	0	100
178	18	129	-	13	15	129	0	129	0	100

Amplicone	Cromosoma	Dimensione del frammento analizzato[1]	Contenuto genomico dell'amplicone	N. di campioni unici	N. totale di campioni analizzati[2]	N. di identificazioni per campione che è stato possibile effettuare[3]	N. di identificazioni non riuscite[4]	N. di identificazioni corrette per campione[5]	N. di identificazioni errate[6]	% di identificazioni corrette[7]
179	18	133	-	13	15	133	0	133	0	100
180	18	118	-	13	15	118	0	118	0	100
181	18	114	60% GC	13	15	114	0	114	0	100
182	18	118	-	13	15	118	0	118	0	100
183	19	122	Poly G (6); 66% GC	13	15	122	0	122	0	100
184	19	139	64% GC	13	15	139	0	139	0	100
185	19	131	67% GC	13	15	131	0	131	0	100
186	19	141	59% GC; Regioni omologhe su un cromosoma diverso	13	15	141	0	141	0	100
187	19	121	Poly C (5); 72% GC; Regioni omologhe su un cromosoma diverso	13	15	121	0	121	0	100
188	19	138	58% GC	13	15	138	0	138	0	100
189	19	123	64% GC	13	15	123	0	123	0	100
190	19	138	-	13	15	138	0	138	0	100

Ampli-cone	Cromo-soma	Dimensione del frammento analizzato[1]	Contenuto genomico dell'amplicone	N. di cam-pioni unici	N. totale di cam-pioni analizzati[2]	N. di identificazioni per campione che è stato possibile effettuare[3]	N. di identifi-cazioni non riuscite[4]	N. di identifi-cazioni corrette per campione[5]	N. di identifi-cazioni errate[6]	% di identificazioni corrette[7]
191	20	117	Poly T (5)	13	15	117	0	117	0	100
192	22	136	Poly A (7)	13	15	136	0	136	0	100
193	22	122	Poly A (5); Poly C (5)	13	15	122	0	122	0	100
194	22	122	62% GC; SNV	13	15	122	0	122	0	100
195	22	119	66% GC	13	15	119	0	119	0	100

[1] Frammenti analizzati indica la dimensione della regione genomica sequenziata in basi, non sono inclusi i primer specifici per il target.

[2] N. totale dei campioni elencati è 15 perché due dei 13 campioni unici analizzati sono stati analizzati in due replicati indipendenti.

[3] N. identificazioni/campione che è stato possibile effettuate rappresenta il numero di basi che presentavano una qualità adeguata per essere identificate dal sistema.

[4] N. di identificazioni non riuscite rappresenta il numero di basi in un amplicone che non sono state identificate nella corsa.

[5] N. di identificazioni corrette per campione rappresenta il numero di basi nell'amplicone che sono state individuate e che hanno fornito risultati che corrispondevano alla sequenza di riferimento del genoma umano versione 19 e al riferimento composito ben caratterizzato.

[6] N. di identificazioni errate rappresenta il numero totale di identificazioni non corrette per SNV o Indel in quell'amplicone; le note a piè di pagina sottostanti forniscono ulteriori dettagli sulle identificazioni non corrette.

[7] La % di identificazioni corrette è pari alla percentuale di identificazione corretta per tutte le basi nell'amplicone, dove l'identificazione corretta per SNV o Indel si basa sulle informazioni dei riferimenti compositi ben caratterizzati e l'identificazione corretta per le basi nella restante sequenza dell'amplicone si basa sul confronto con la sequenza di riferimento del genoma umano versione 19. Questa colonna potrebbe presentare più di un risultato previsto per un dato amplicone se alcuni campioni contengono un Indel o no, ad esempio, amplicone 9. La % di identificazioni corrette per i campioni con un risultato non corretto è presentata nella tabella.

[8] L'amplicone 9 presenta una corsa per omopolimeri di 14 A in base alla sequenza di riferimento del genoma umano versione 19. Tuttavia, le informazioni per il riferimento composito ben caratterizzato per 7 dei 13 campioni presentano 13 A in questa corsa per omopolimeri. In questi 7 campioni, questa delezione di una coppia di basi rappresenta un falso negativo nello studio di accuratezza del sequenziamento MiSeqDx.

[9] L'amplicone 46 presenta un'inserzione di una base che viene riportata nei 9 campioni nel database dei riferimenti ben caratterizzati e viene rilevata correttamente in tutti i campioni analizzati.

[10] L'amplicone 95 presenta una corsa per omopolimeri di 14 A in base alla sequenza di riferimento del genoma umano versione 19. Tuttavia, le sequenze per il riferimento composito ben caratterizzato per 13 dei 13 campioni presentano 15 A in questa corsa per omopolimeri. In questi 13 campioni, questa inserzione di una coppia di basi rappresenta un falso negativo nello studio di accuratezza del sequenziamento MiSeqDx.

La tabella seguente contiene i dati dallo Studio 1 presentati con una percentuale di concordanza positiva e negativa, dove i risultati delle varianti sono confrontati sulle informazioni dei riferimenti compositi ben caratterizzati per i calcoli PPA. Poiché le informazioni dei riferimenti compositi forniscono solo risultati per le varianti di singolo nucleotide e inserzioni/delezioni, i risultati delle basi non varianti sono confrontati alla sequenza di riferimento del genoma umano versione 19, per i calcoli NPA. Tutte le basi non varianti presentano una concordanza del 100% con la sequenza di riferimento. Tutte le SNV presentano una concordanza del 100% con la sequenza di riferimento. Le varianti che non sono state identificate erano inserzioni di 1 base o delezioni di 1 base nelle regioni omopolimeriche.

Tabella 2 Concordanza dei risultati di identificazione delle basi sulla piattaforma MiSeqDx con i dati di riferimento per 13 campioni ben caratterizzati

Campione	N. di ampliconi	% di copertura degli ampliconi[1]	Varianti previste per campione[2]	Varianti identificate correttamente	Varianti non identificate[3]	Basi non varianti identificate correttamente	PPA[4] (%)	NPA[5] (%)
NA12877	195	100	19	17	2	24418	89,47	100
NA12878	195	100	19	17	2	24417	89,47	100
NA12879	195	100	20	19	1	24416	95,00	100
NA12880	195	100	20	18	2	24417	90,00	100
NA12881	195	100	22	20	2	24415	90,91	100
NA12882	195	100	16	15	1	24419	93,75	100
NA12883	195	100	24	23	1	24412	95,83	100
NA12884	195	100	21	20	1	24415	95,24	100
NA12885	195	100	19	17	2	24417	89,47	100
NA12886	195	100	22	20	2	24415	90,91	100
NA12887	195	100	19	18	1	24416	94,74	100
NA12888	195	100	24	23	1	24412	95,83	100
NA12893	195	100	20	18	2	24417	90,00	100

[1] % di copertura degli ampliconi rappresenta il numero di basi negli ampliconi sequenziate con affidabilità.
[2] Varianti previste per campione include sia le SNV che gli Indel.
[3] Per le varianti non identificate, vedere la prima tabella per lo Studio 1 e le note a piè di pagina 8-10.
[4] Percentuale di concordanza positiva (PPA) = 100xTP/(TP+FN) dove i veri positivi (TP) rappresentano il numero di identificazioni delle varianti positive a coordinate genomiche in cui le varianti sono presenti in base alla sequenza di riferimento e l'allele mutante identificato concorda con la sequenza di riferimento (colonna denominata "Varianti identificate correttamente") e i falsi negativi (FN) rappresentano il numero di identificazioni delle varianti negative a coordinate genomiche in cui le varianti sono presenti in base alla sequenza di riferimento (colonna denominata "Varianti identificate").
[5] Percentuale di concordanza negativa (NPA) = 100xTN/(FP+TN) dove i falsi positivi (FP) rappresentano il numero di identificazioni delle varianti positive a coordinate genomiche in cui le varianti sono assenti in base alla sequenza di riferimento o se l'allele mutante identificato è discordante con la sequenza di riferimento (non nella tabella; in questo studio non sono state eseguite identificazioni delle varianti falso positive) e veri negativi (TN) rappresentano il numero di identificazioni delle varianti negative a coordinate genomiche in cui le varianti sono assenti in base allo standard di riferimento (colonna denominata "Basi non varianti identificate correttamente").

Studio 2

I risultati del sequenziamento per il pannello degli ampliconi sopra indicato sono stati confrontati con un genotipo altamente affidabile per NA12878 dal National Institutes of Standards and Technology (NIST) (v.2.15[1]). Dei 195 ampliconi, 184 ampliconi rientravano nelle identificazioni di riferimento altamente affidabili nella sequenza NIST ed erano stati confrontati. Le identificazioni di basi non varianti sono state confrontate con la sequenza di riferimento del genoma umano versione 19.

Tabella 3 Confronto dei risultati del sequenziamento sulla piattaforma MiSeqDx per il campione NA12878 con il database NIST

Campione	N. di ampliconi	% di copertura degli ampliconi[2]	Varianti previste	Varianti identificate correttamente	Varianti non identificate	Basi non varianti identificate correttamente	PPA[3] (%)	NPA[4] (%)
NA12878	184	100	17	16	1[5]	23066	94,12	100

[1] Zook, JM et al. Integrating sequencing datasets to form highly confident SNP and indel genotype calls for a whole human genome. arXiv:1307.4661 [q-bio.GN].
[2] % di copertura degli ampliconi rappresenta il numero di basi negli ampliconi sequenziate con affidabilità.
[3] Percentuale di concordanza positiva (PPA) = 100xTP/(TP+FN).
[4] Percentuale di concordanza negativa (NPA) = 100xTN/(FP+TN)
[5] La variante non identificata rappresenta la delezione di una coppia di basi nell'amplicone 9 nella corsa di omopolimeri di 14 A non identificata da MiSeqDx che è presente nella sequenza NIST. Prestare attenzione che la sequenza NIST non include l'inserzione di una coppia di basi nell'altro omopolimero di A che era presente nell'altro database di riferimento usato sopra nello Studio 1.

Studio 3

È stato eseguito un ulteriore studio di accuratezza per valutare le prestazioni di inserzioni e delezioni piccole in un saggio rappresentativo, il saggio MiSeqDx Cystic Fibrosis 139-Variant Illumina, che include un sottogruppo di variazioni genetiche del gene *CFTR* significative dal punto di vista clinico e analizzato con il software MiSeq Reporter usando il flusso di lavoro di sequenziamento del DNA target sulla piattaforma MiSeqDx. Le inserzioni e le delezioni interrogate sono state rilevate dove previsto con elevata affidabilità. Questi campioni sono stati caratterizzati mediante il sequenziamento bidirezionale Sanger, un metodo di riferimento usato per stabilire le sequenze previste.

Tabella 4 Riepilogo delle delezioni Indel con la piattaforma MiSeqDx

Amplicone	Dimensione inserto	Contenuto genomico dell'amplicone	N. di identificazioni/campione che è stato possibile effettuare	N. di basi identificate/campione	N. di mancate identificazioni	N. di identificazioni/campione corrette	N. di identificazioni errate	% di identificazioni corrette
1	129	Inserzione di 1 base	130	130	0	130	0	100
2	154	Delezione di 3 basi	151	151	0	151	0	100
3	167	Delezione di 2 basi	165	165	0	165	0	100

Amplicone	Dimensione inserto	Contenuto genomico dell'amplicone	N. di identificazioni/ campione che è stato possibile effettuare	N. di basi identificate/ campione	N. di mancate identificazioni	N. di identificazioni/ campione corrette	N. di identificazioni errate	% di identificazioni corrette
4	134	Delezione di 1 base	133	133	0	133	0	100
5	132	Delezione di 1 base	131	131	0	131	0	100
6	129	Delezione di 1 base	128	128	0	128	0	100

I dati forniti da questi studi di accuratezza supportano l'asserzione che la piattaforma MiSeqDx può sequenziare accuratamente:

- Contenuto GC ≥ 19% (tutte le basi in 135 dei 135 ampliconi sequenziati con 19% di contenuto GC identificato correttamente)
- Contenuto GC ≤ 72% (tutte le basi in 135 dei 135 ampliconi sequenziati con 72% di contenuto GC identificato correttamente)
- Lunghezze PolyA ≤ 7 (ripetizione PolyA di 7 nucleotidi era stata identificata correttamente in 270 dei 270 ampliconi sequenziati contenenti PolyA =7)
- Lunghezze PolyT ≤ 8 (ripetizione PolyT di 8 nucleotidi era stata identificata correttamente in 270 dei 270 ampliconi sequenziati contenenti PolyT =8)
- Lunghezze PolyG ≤ 6 (ripetizione PolyG di 6 nucleotidi era stata identificata correttamente in 405 dei 405 ampliconi sequenziati contenenti PolyG =6)
- Lunghezze PolyC ≤ 7 (ripetizione PolyC di 7 nucleotidi era stata identificata correttamente in 135 dei 135 ampliconi sequenziati contenenti PolyC =7)
- Lunghezze di dinucleotidi ripetuti ≤ 5x (tutte le basi in 135 dei 135 ampliconi sequenziati con 5x di ripetizione di dinucleotidi sono state identificate correttamente)
- Lunghezze di trinucleotidi ripetuti ≤ 4x (tutte le basi in 810 dei 810 ampliconi sequenziati con 4x di ripetizione di trinucleotidi sono state identificate correttamente)
- Inserzioni di 1 base e delezioni di 3 o meno basi
  - 2 delle 3 inserzioni di 1 base testati sono state identificate correttamente. Le identificazioni corrette sono state fatte per due inserzioni di 1 base in regioni non omopolimeriche in 82 ampliconi. Un'inserzione di 1 base non era stata identificata in una corsa per omopolimeri di 14 A su cromosoma 2 in 135 ampliconi.
  - 3 delle 4 delezioni di 1 base sono state identificate correttamente. Tutte le identificazioni corrette sono state eseguite in regioni non omopolimeriche in 4 ampliconi. Una delezione di 1 base non era stata identificata in una corsa per omopolimeri di 14 A su cromosoma 9 in 63 ampliconi.
  - Le delezioni di 2 basi sono state identificate correttamente in un campione.
  - Le delezioni di 3 basi sono state identificate correttamente in 21 campioni.

Riproducibilità

La riproducibilità della piattaforma MiSeqDx è stata determinata usando due saggi rappresentativi.

Studio 1

È stato progettato un saggio rappresentativo per interrogare una varietà di geni che coprono 24.434 basi su 19 diversi cromosomi e che contengono potenzialmente gli esoni rilevanti dal punto di vista clinico. Lo studio ha esaminato 13 campioni su nove corse usando tre diversi strumenti MiSeqDx e tre diversi operatori (Tabella 5). Sono stati usati un singolo lotto di reagenti per la preparazione della libreria e due lotti di materiali di consumo per il sequenziamento. I 13 campioni erano stati prelevati da due genitori e 11 bambini. Questi campioni sono spesso stati sottoposti a sequenziamento da laboratori multipli e metodi di sequenziamento multipli. Due campioni sono stati analizzati in duplicati, in questo modo ciascuna corsa ha generato risultati per 15 campioni.

Per la valutazione della riproducibilità da lotto a lotto, sono stati analizzati 94 campioni e due controlli non templati su tre lotti. Ciascun lotto è stato diviso in corse da 48 campioni per testare tutti i reagenti e le possibili combinazioni di primer indice. Tutte le corse di sequenziamento sono state completate da un singolo operatore e su un singolo strumento MiSeqDx per rimuovere qualsiasi potenziale variazione da parte dell'operatore o dello strumento (Tabella 6).

Le identificazioni corrette sono state determinate per le varianti di singolo nucleotide (SNV) confrontando i dati dello studio su informazioni dei riferimenti ben caratterizzati. Per lo studio di riproducibilità non si sono verificate corse non riuscite o corse ripetute. Le tabelle seguenti mostrano i risultati degli studi per valutare la riproducibilità del sistema.

Tabella 5 Studio 1 - Risultati sulla riproducibilità da strumento a strumento per la piattaforma MiSeqDx (a livello di ampliconi)

Amplicone	Crom.	Dimensione del frammento analizzato[1]	Contenuto genomico dell'amplicone	N. di campioni analizzati[2]	MiSeqDx 1			MiSeqDx 2			MiSeqDx 3		
					N. totale di identificazioni non riuscite[3]	N. totale di identificazioni errate[4]	% di identificazioni corrette[5]	N. totale di identificazioni non riuscite[3]	N. totale di identificazioni errate[4]	% di identificazioni corrette[5]	N. totale di identificazioni non riuscite[3]	N. totale di identificazioni errate[4]	% di identificazioni corrette[5]
1	1	132	Poly C (5); 63% GC	135	0	0	100	23[6]	0	99,61[7]	39[6]	0	99,34[7]
2	1	128	Poly T (5)	135	0	0	100	0	0	100	0	0	100
3	2	133	-	135	0	0	100	0	0	100	0	0	100
4	2	119	-	135	0	0	100	0	0	100	0	0	100
5	2	127	Poly T (5)	135	0	0	100	0	0	100	0	0	100
6	2	135	Poly A (6)	135	0	0	100	0	0	100	0	0	100
7	2	122	Poly T (5); Poly C (5)	135	0	0	100	0	0	100	0	0	100
8	2	110	Poly T (5)	135	0	0	100	0	0	100	0	0	100
9	2	131	Poly A (14)	135	0	27[8]	99,54	0	27[8]	99,54	0	27[8]	99,54
10	2	117	-	135	0	0	100	0	0	100	0	0	100
11	2	121	-	135	0	0	100	0	0	100	0	0	100
12	2	114	-	135	0	0	100	0	0	100	0	0	100
13	2	129	Poly A (5)	135	0	0	100	0	0	100	0	0	100

Amplicone	Crom.	Dimensione del frammento analizzato[1]	Contenuto genomico dell'amplicone	N. di campioni analizzati[2]	MiSeqDx 1			MiSeqDx 2			MiSeqDx 3		
					N. totale di identificazioni non riuscite[3]	N. totale di identificazioni errate[4]	% di identificazioni corrette[5]	N. totale di identificazioni non riuscite[3]	N. totale di identificazioni errate[4]	% di identificazioni corrette[5]	N. totale di identificazioni non riuscite[3]	N. totale di identificazioni errate[4]	% di identificazioni corrette[5]
14	3	131	Poly A (5); Poly T (5)	135	0	0	100	0	0	100	0	0	100
15	3	130	-	135	0	0	100	0	0	100	0	0	100
16	3	130	-	135	0	0	100	0	0	100	0	0	100
17	3	117	-	135	0	0	100	0	0	100	0	0	100
18	3	136	Poly T (5)	135	0	0	100	0	0	100	0	0	100
19	3	131	Poly T (5); SNV	135	0	0	100	0	0	100	0	0	100
20	3	123	Poly A (5)	135	0	0	100	0	0	100	0	0	100
21	3	117	Poly A (6); Poly T (5); Regioni omologhe su un cromosoma diverso	135	0	0	100	0	0	100	0	0	100
22	3	119	Regioni omologhe su un cromosoma diverso	135	0	0	100	0	0	100	0	0	100

Amplicone	Crom.	Dimensione del frammento analizzato[1]	Contenuto genomico dell'amplicone	N. di campioni analizzati[2]	MiSeqDx 1			MiSeqDx 2			MiSeqDx 3		
					N. totale di identificazioni non riuscite[3]	N. totale di identificazioni errate[4]	% di identificazioni corrette[5]	N. totale di identificazioni non riuscite[3]	N. totale di identificazioni errate[4]	% di identificazioni corrette[5]	N. totale di identificazioni non riuscite[3]	N. totale di identificazioni errate[4]	% di identificazioni corrette[5]
23	3	120	-	135	0	0	100	0	0	100	0	0	100
24	3	129	Poly T (5)	135	0	0	100	0	0	100	0	0	100
25	4	133	Poly C (7); 66% GC	135	0	0	100	0	0	100	0	0	100
26	4	135	Poly C (5); 69% GC	135	0	0	100	0	0	100	0	0	100
27	4	123	SNV	135	0	0	100	0	0	100	0	0	100
28	4	134	-	135	0	0	100	0	0	100	0	0	100
29	4	132	-	135	0	0	100	0	0	100	0	0	100
30	4	121	Poly A (5); SNV	135	0	0	100	0	0	100	0	0	100
31	4	125	-	135	0	0	100	0	0	100	0	0	100
32	4	134	Poly T (5)	135	0	0	100	0	0	100	0	0	100
33	4	118	-	135	0	0	100	0	0	100	0	0	100
34	4	122	Poly A (5)	135	0	0	100	0	0	100	0	0	100
35	4	131	-	135	0	0	100	0	0	100	0	0	100
36	4	133	-	135	0	0	100	0	0	100	0	0	100

Amplicone	Crom.	Dimensione del frammento analizzato[1]	Contenuto genomico dell'amplicone	N. di campioni analizzati[2]	MiSeqDx 1			MiSeqDx 2			MiSeqDx 3		
					N. totale di identificazioni non riuscite[3]	N. totale di identificazioni errate[4]	% di identificazioni corrette[5]	N. totale di identificazioni non riuscite[3]	N. totale di identificazioni errate[4]	% di identificazioni corrette[5]	N. totale di identificazioni non riuscite[3]	N. totale di identificazioni errate[4]	% di identificazioni corrette[5]
37	4	128	Poly T (6)	135	0	0	100	0	0	100	0	0	100
38	4	131	-	135	0	0	100	0	0	100	0	0	100
39	4	129	Poly A (5); Poly T (5); SNV	135	0	0	100	0	0	100	0	0	100
40	4	133	Poly T (5); SNV	135	0	0	100	0	0	100	0	0	100
41	4	112	SNV	135	0	0	100	0	0	100	0	0	100
42	4	133	-	135	0	0	100	0	0	100	0	0	100
43	4	135	-	135	0	0	100	0	0	100	0	0	100
44	4	122	-	135	0	0	100	0	0	100	0	0	100
45	4	117	-	135	0	0	100	0	0	100	0	0	100
46	4	124	-	135	0	0	100	0	0	100	0	0	100
47	4	117	Poly T (5)	135	0	0	100	0	0	100	0	0	100
48	4	128	Poly A (7)	135	0	0	100	0	0	100	0	0	100
49	4	123	Poly A (6)	135	0	0	100	0	0	100	0	0	100
50	4	133	-	135	0	0	100	0	0	100	0	0	100

Amplicone	Crom.	Dimensione del frammento analizzato[1]	Contenuto genomico dell'amplicone	N. di campioni analizzati[2]	MiSeqDx 1			MiSeqDx 2			MiSeqDx 3		
					N. totale di identificazioni non riuscite[3]	N. totale di identificazioni errate[4]	% di identificazioni corrette[5]	N. totale di identificazioni non riuscite[3]	N. totale di identificazioni errate[4]	% di identificazioni corrette[5]	N. totale di identificazioni non riuscite[3]	N. totale di identificazioni errate[4]	% di identificazioni corrette[5]
51	4	112	-	135	0	0	100	0	0	100	0	0	100
52	4	129	-	135	0	0	100	0	0	100	0	0	100
53	4	126	-	135	0	0	100	0	0	100	0	0	100
54	4	132	-	135	0	0	100	0	0	100	0	0	100
55	5	131	-	135	0	0	100	0	0	100	0	0	100
56	5	119	-	135	0	0	100	0	0	100	0	0	100
57	5	120	Poly A (5)	135	0	0	100	0	0	100	0	0	100
58	5	119	-	135	0	0	100	0	0	100	0	0	100
59	5	118	-	135	0	0	100	0	0	100	0	0	100
60	5	112	-	135	0	0	100	0	0	100	0	0	100
61	5	120	-	135	0	0	100	0	0	100	0	0	100
62	5	120	Poly A (5)	135	0	0	100	0	0	100	0	0	100
63	5	115	CT(5)	135	0	0	100	0	0	100	0	0	100
64	5	112	SNV	135	0	0	100	0	0	100	0	0	100
65	5	135	Poly T (6)	135	0	0	100	0	0	100	0	0	100
66	5	131	63% GC	135	0	0	100	0	0	100	0	0	100

Amplicone	Crom.	Dimensione del frammento analizzato[1]	Contenuto genomico dell'amplicone	N. di campioni analizzati[2]	MiSeqDx 1			MiSeqDx 2			MiSeqDx 3		
					N. totale di identificazioni non riuscite[3]	N. totale di identificazioni errate[4]	% di identificazioni corrette[5]	N. totale di identificazioni non riuscite[3]	N. totale di identificazioni errate[4]	% di identificazioni corrette[5]	N. totale di identificazioni non riuscite[3]	N. totale di identificazioni errate[4]	% di identificazioni corrette[5]
67	5	121	-	135	0	0	100	0	0	100	0	0	100
68	5	132	Poly A (6); Poly T (8)	135	0	0	100	0	0	100	0	0	100
69	7	133	-	135	0	0	100	0	0	100	0	0	100
70	7	120	60% GC	135	0	0	100	0	0	100	0	0	100
71	7	135	-	135	0	0	100	0	0	100	0	0	100
72	7	126	Poly A (5); 59% GC	135	0	0	100	0	0	100	0	0	100
73	7	134	-	135	0	0	100	0	0	100	0	0	100
74	7	122	Poly C (5); 63% GC	135	0	0	100	0	0	100	0	0	100
75	7	127	59% GC; SNV	135	0	0	100	0	0	100	0	0	100
76	7	123	-	135	0	0	100	0	0	100	0	0	100
77	7	125	-	135	0	0	100	0	0	100	0	0	100
78	7	133	Poly A (5); Poly T (5)	135	0	0	100	0	0	100	0	0	100

Amplicone	Crom.	Dimensione del frammento analizzato[1]	Contenuto genomico dell'amplicone	N. di campioni analizzati[2]	MiSeqDx 1			MiSeqDx 2			MiSeqDx 3		
					N. totale di identificazioni non riuscite[3]	N. totale di identificazioni errate[4]	% di identificazioni corrette[5]	N. totale di identificazioni non riuscite[3]	N. totale di identificazioni errate[4]	% di identificazioni corrette[5]	N. totale di identificazioni non riuscite[3]	N. totale di identificazioni errate[4]	% di identificazioni corrette[5]
79	7	116	-	135	0	0	100	0	0	100	0	0	100
80	7	135	-	135	0	0	100	0	0	100	0	0	100
81	7	118	-	135	0	0	100	0	0	100	0	0	100
82	7	136	67% GC	135	0	0	100	0	0	100	0	0	100
83	7	131	58% GC	135	0	0	100	0	0	100	0	0	100
84	7	119	Poly G (6); 61% GC	135	0	0	100	0	0	100	0	0	100
85	7	122	Poly T (5)	135	0	0	100	0	0	100	0	0	100
86	7	123	Poly A (6)	135	0	0	100	0	0	100	0	0	100
87	8	127	60% GC	135	0	0	100	0	0	100	0	0	100
88	8	129	57% GC	135	0	0	100	0	0	100	0	0	100
89	9	130	Poly T (5)	135	0	0	100	0	0	100	0	0	100
90	9	116	-	135	0	0	100	0	0	100	0	0	100
91	9	119	Regioni omologhe su un cromosoma diverso	135	0	0	100	0	0	100	0	0	100

Amplicone	Crom.	Dimensione del frammento analizzato[1]	Contenuto genomico dell'amplicone	N. di campioni analizzati[2]	MiSeqDx 1			MiSeqDx 2			MiSeqDx 3		
					N. totale di identificazioni non riuscite[3]	N. totale di identificazioni errate[4]	% di identificazioni corrette[5]	N. totale di identificazioni non riuscite[3]	N. totale di identificazioni errate[4]	% di identificazioni corrette[5]	N. totale di identificazioni non riuscite[3]	N. totale di identificazioni errate[4]	% di identificazioni corrette[5]
92	9	121	-	135	0	0	100	0	0	100	0	0	100
93	9	117	Regioni omologhe su un cromosoma diverso	135	0	0	100	0	0	100	0	0	100
94	9	114	-	135	0	0	100	0	0	100	0	0	100
95	9	129	Poly A (14)	135	0	45[9]	99,22	0	45[9]	99,22	0	45[9]	99,22
96	9	114	Regioni omologhe su un cromosoma diverso; SNV	135	0	0	100	0	0	100	0	0	100
97	9	122	-	135	0	0	100	0	0	100	0	0	100
98	9	127	Poly A (5); Poly C (5)	135	0	0	100	0	0	100	0	0	100
99	9	133	-	135	0	0	100	0	0	100	0	0	100
100	9	138	64% GC	135	0	0	100	0	0	100	0	0	100
101	9	139	-	135	0	0	100	0	0	100	0	0	100

Amplicone	Crom.	Dimensione del frammento analizzato[1]	Contenuto genomico dell'amplicone	N. di campioni analizzati[2]	MiSeqDx 1			MiSeqDx 2			MiSeqDx 3		
					N. totale di identificazioni non riuscite[3]	N. totale di identificazioni errate[4]	% di identificazioni corrette[5]	N. totale di identificazioni non riuscite[3]	N. totale di identificazioni errate[4]	% di identificazioni corrette[5]	N. totale di identificazioni non riuscite[3]	N. totale di identificazioni errate[4]	% di identificazioni corrette[5]
102	9	116	-	135	0	0	100	0	0	100	0	0	100
103	9	133	Poly A (5); 57% GC	135	0	0	100	0	0	100	0	0	100
104	9	138	57% GC	135	0	0	100	0	0	100	0	0	100
105	9	136	Poly C (5); 67% GC	135	0	0	100	0	0	100	0	0	100
106	9	118	70% GC	135	0	0	100	0	0	100	0	0	100
107	10	128	62% GC	135	0	0	100	0	0	100	0	0	100
108	10	120	60% GC	135	0	0	100	0	0	100	0	0	100
109	10	139	58% GC; SNV	135	0	0	100	0	0	100	0	0	100
110	10	118	57% GC	135	0	0	100	0	0	100	0	0	100
111	10	123	Poly T (5)	135	0	0	100	0	0	100	0	0	100
112	10	121	-	135	0	0	100	0	0	100	0	0	100
113	10	129	26% GC	135	0	0	100	0	0	100	0	0	100
114	10	122	-	135	0	0	100	0	0	100	0	0	100

Ampli-cone	Crom.	Dimensione del frammento analizzato[1]	Contenuto genomico dell'amplicone	N. di campioni analizzati[2]	MiSeqDx 1			MiSeqDx 2			MiSeqDx 3		
					N. totale di identificazioni non riuscite[3]	N. totale di identificazioni errate[4]	% di identificazioni corrette[5]	N. totale di identificazioni non riuscite[3]	N. totale di identificazioni errate[4]	% di identificazioni corrette[5]	N. totale di identificazioni non riuscite[3]	N. totale di identificazioni errate[4]	% di identificazioni corrette[5]
115	10	124	Poly T (5); Regioni omologhe su un cromosoma diverso	135	0	0	100	0	0	100	0	0	100
116	10	135	CA(4)	135	0	0	100	0	0	100	0	0	100
117	10	135	Poly A (6); Regioni omologhe su un cromosoma diverso	135	0	0	100	0	0	100	0	0	100
118	10	119	Poly C (5); SNV	135	0	0	100	0	0	100	0	0	100
119	10	125	-	135	0	0	100	0	0	100	0	0	100
120	10	131	-	135	0	0	100	0	0	100	0	0	100
121	10	117	-	135	0	0	100	0	0	100	0	0	100
122	10	116	-	135	0	0	100	0	0	100	0	0	100
123	10	129	58% GC	135	0	0	100	0	0	100	0	0	100

Ampli-cone	Crom.	Dimen-sione del fram-mento analiz-zato[1]	Contenuto genomico dell'am-plicone	N. di cam-pioni analiz-zati[2]	MiSeqDx 1			MiSeqDx 2			MiSeqDx 3		
					N. totale di identi-ficazio-ni non riuscite[3]	N. totale di identi-ficazio-ni errate[4]	% di identi-ficazio-ni corrette[5]	N. totale di identi-ficazio-ni non riuscite[3]	N. totale di identi-ficazio-ni errate[4]	% di identi-ficazio-ni corrette[5]	N. totale di identi-ficazio-ni non riuscite[3]	N. totale di identi-ficazio-ni errate[4]	% di identi-ficazio-ni corrette[5]
124	11	117	Poly T (10)	135	0	0	100	0	0	100	0	0	100
125	11	117	Poly T (5)	135	0	0	100	0	0	100	0	0	100
126	11	113	Poly A (5)	135	0	0	100	0	0	100	0	0	100
127	11	129	-	135	0	0	100	0	0	100	0	0	100
128	11	121	Poly T (5)	135	0	0	100	0	0	100	0	0	100
129	11	123	-	135	0	0	100	0	0	100	0	0	100
130	11	127	Poly A (6)	135	0	0	100	0	0	100	0	0	100
131	11	136	Poly T (6)	135	0	0	100	0	0	100	0	0	100
132	11	132	Poly T (5)	135	0	0	100	0	0	100	0	0	100
133	11	115	-	135	0	0	100	0	0	100	0	0	100
134	11	117	Poly T (8); 19% GC	135	0	0	100	0	0	100	0	0	100
135	11	134	Poly A (5); Poly T (5)	135	0	0	100	0	0	100	0	0	100
136	11	131	Poly A (5)	135	0	0	100	0	0	100	0	0	100
137	11	133	SNV; 26% GC	135	0	0	100	0	0	100	0	0	100

Ampli-cone	Crom.	Dimensione del frammento analizzato[1]	Contenuto genomico dell'amplicone	N. di campioni analizzati[2]	MiSeqDx 1			MiSeqDx 2			MiSeqDx 3		
					N. totale di identificazioni non riuscite[3]	N. totale di identificazioni errate[4]	% di identificazioni corrette[5]	N. totale di identificazioni non riuscite[3]	N. totale di identificazioni errate[4]	% di identificazioni corrette[5]	N. totale di identificazioni non riuscite[3]	N. totale di identificazioni errate[4]	% di identificazioni corrette[5]
138	11	137	Poly T (8); SNV	135	0	0	100	0	0	100	0	0	100
139	11	131	Poly A (5)	135	0	0	100	0	0	100	0	0	100
140	12	131	-	135	0	0	100	0	0	100	0	0	100
141	12	128	-	135	0	0	100	0	0	100	0	0	100
142	12	133	Poly A (5)	135	0	0	100	0	0	100	0	0	100
143	12	136	-	135	0	0	100	0	0	100	0	0	100
144	12	124	-	135	0	0	100	0	0	100	0	0	100
145	12	122	59% GC	135	0	0	100	0	0	100	0	0	100
146	13	122	-	135	0	0	100	0	0	100	0	0	100
147	13	116	Poly C (5)	135	0	0	100	0	0	100	0	0	100
148	13	133	-	135	0	0	100	0	0	100	0	0	100
149	13	117	SNV	135	0	0	100	0	0	100	0	0	100
150	13	124	Poly T (6)	135	0	0	100	0	0	100	0	0	100
151	13	123	Poly T (5); 26% GC	135	0	0	100	0	0	100	0	0	100

Amplicone	Crom.	Dimensione del frammento analizzato[1]	Contenuto genomico dell'amplicone	N. di campioni analizzati[2]	MiSeqDx 1			MiSeqDx 2			MiSeqDx 3		
					N. totale di identificazioni non riuscite[3]	N. totale di identificazioni errate[4]	% di identificazioni corrette[5]	N. totale di identificazioni non riuscite[3]	N. totale di identificazioni errate[4]	% di identificazioni corrette[5]	N. totale di identificazioni non riuscite[3]	N. totale di identificazioni errate[4]	% di identificazioni corrette[5]
152	13	115	Poly A (5)	135	0	0	100	0	0	100	0	0	100
153	13	125	-	135	0	0	100	0	0	100	0	0	100
154	13	121	-	135	0	0	100	0	0	100	0	0	100
155	13	123	-	135	0	0	100	0	0	100	0	0	100
156	13	114	-	135	0	0	100	0	0	100	0	0	100
157	13	119	-	135	0	0	100	0	0	100	0	0	100
158	14	122	58% GC	135	0	0	100	0	0	100	0	0	100
159	16	122	-	135	0	0	100	0	0	100	0	0	100
160	16	121	-	135	0	0	100	0	0	100	0	0	100
161	16	123	Poly C (5)	135	0	0	100	0	0	100	0	0	100
162	17	119	-	135	0	0	100	0	0	100	0	0	100
163	17	119	61% GC	135	0	0	100	0	0	100	0	0	100
164	17	135	-	135	0	0	100	0	0	100	0	0	100
165	17	116	Poly C (6); 60% GC; SNV	135	0	0	100	0	0	100	0	0	100

Ampli-cone	Crom.	Dimen-sione del fram-mento analiz-zato[1]	Contenuto genomico dell'am-plicone	N. di cam-pioni analiz-zati[2]	MiSeqDx 1			MiSeqDx 2			MiSeqDx 3		
					N. totale di identi-ficazio-ni non riuscite[3]	N. totale di identi-ficazio-ni errate[4]	% di identi-ficazio-ni corrette[5]	N. totale di identi-ficazio-ni non riuscite[3]	N. totale di identi-ficazio-ni errate[4]	% di identi-ficazio-ni corrette[5]	N. totale di identi-ficazio-ni non riuscite[3]	N. totale di identi-ficazio-ni errate[4]	% di identi-ficazio-ni corrette[5]
166	17	123	-	135	0	0	100	0	0	100	0	0	100
167	17	116	62% GC	135	0	0	100	0	0	100	0	0	100
168	17	118	Poly C (5); 65% GC	135	0	0	100	0	0	100	0	0	100
169	17	129	-	135	0	0	100	0	0	100	0	0	100
170	17	131	Poly G (6); 67% GC; SNV	135	0	0	100	0	0	100	0	0	100
171	17	127	61% GC	135	0	0	100	0	0	100	0	0	100
172	17	118	Poly C (5)	135	0	0	100	0	0	100	0	0	100
173	17	138	61% GC	135	0	0	100	0	0	100	0	0	100
174	17	131	58% GC	135	0	0	100	0	0	100	0	0	100
175	18	112	-	135	0	0	100	0	0	100	0	0	100
176	18	124	-	135	0	0	100	0	0	100	0	0	100
177	18	134	Poly A (6)	135	0	0	100	0	0	100	0	0	100
178	18	129	-	135	0	0	100	0	0	100	0	0	100
179	18	133	-	135	0	0	100	0	0	100	0	0	100

Amplicone	Crom.	Dimensione del frammento analizzato[1]	Contenuto genomico dell'amplicone	N. di campioni analizzati[2]	MiSeqDx 1			MiSeqDx 2			MiSeqDx 3		
					N. totale di identificazioni non riuscite[3]	N. totale di identificazioni errate[4]	% di identificazioni corrette[5]	N. totale di identificazioni non riuscite[3]	N. totale di identificazioni errate[4]	% di identificazioni corrette[5]	N. totale di identificazioni non riuscite[3]	N. totale di identificazioni errate[4]	% di identificazioni corrette[5]
180	18	118	-	135	0	0	100	0	0	100	0	0	100
181	18	114	60% GC	135	0	0	100	0	0	100	0	0	100
182	18	118	-	135	0	0	100	0	0	100	0	0	100
183	19	122	Poly G (6); 66% GC	135	0	0	100	0	0	100	0	0	100
184	19	139	64% GC	135	0	0	100	0	0	100	0	0	100
185	19	131	67% GC	135	0	0	100	0	0	100	0	0	100
186	19	141	59% GC; Regioni omologhe su un cromosoma diverso	135	0	0	100	0	0	100	0	0	100
187	19	121	Poly C (5); 72% GC; Regioni omologhe su un cromosoma diverso	135	0	0	100	0	0	100	0	0	100
188	19	138	58% GC	135	0	0	100	0	0	100	0	0	100

Amplicone	Crom.	Dimensione del frammento analizzato[1]	Contenuto genomico dell'amplicone	N. di campioni analizzati[2]	MiSeqDx 1			MiSeqDx 2			MiSeqDx 3		
					N. totale di identificazioni non riuscite[3]	N. totale di identificazioni errate[4]	% di identificazioni corrette[5]	N. totale di identificazioni non riuscite[3]	N. totale di identificazioni errate[4]	% di identificazioni corrette[5]	N. totale di identificazioni non riuscite[3]	N. totale di identificazioni errate[4]	% di identificazioni corrette[5]
189	19	123	64% GC	135	0	0	100	0	0	100	0	0	100
190	19	138	-	135	0	0	100	0	0	100	0	0	100
191	20	117	Poly T (5)	135	0	0	100	0	0	100	0	0	100
192	22	136	Poly A (7)	135	0	0	100	0	0	100	0	0	100
193	22	122	Poly A (5); Poly C (5)	135	0	0	100	0	0	100	0	0	100
194	22	122	62% GC; SNV	135	0	0	100	0	0	100	0	0	100
195	22	119	66% GC	135	0	0	100	0	0	100	0	0	100

[1] Frammenti analizzati indica la dimensione della regione genomica sequenziata in basi, non sono inclusi i primer specifici per il target.
[2] Numero di campioni è calcolato da 9 corse di 15 campioni (una volta la corsa da 11 campioni e due volte la corsa da 2 campioni).
[3] Numero totale di identificazioni non riuscite rappresenta il numero combinato di mancate identificazioni ottenute per tutte le 45 corse che analizzavano l'amplicone specifico usando uno strumento MiSeqDx specifico.
[4] Numero totale di identificazioni errate rappresenta il numero combinato di identificazioni errate ottenute per tutte le 45 corse che analizzavano l'amplicone specifico usando uno strumento MiSeqDx specifico.
[5] La % di identificazioni corrette è pari alla percentuale di identificazioni corrette per tutte le basi nell'amplicone, dove l'identificazione corretta per SNV o Indel si basa sul database dei riferimenti compositi ben caratterizzati e l'identificazione corretta per le basi nella restante sequenza dell'amplicone si basa sul confronto con la sequenza di riferimento del genoma umano versione 19. Questa colonna potrebbe presentare più di un risultato previsto per un dato amplicone se per alcuni campioni è prevista la presenza di uno o più Indel, ad esempio, amplicone 9.
[6] L'amplicone 1 presentava diverse basi il cui genotipo non è stato possibile identificare: 12 basi nelle corse 1/9 in NA12881; 1 base nelle corse 2/9 e 3 basi nelle corse 1/9 in NA12886; 20 basi nelle corse 1/9 e 26 basi nelle corse 1/9 in NA12888. Questo è dovuto alla copertura bassa a basi con identificazione non riuscita in quelle corse, dove la profondità di sequenziamento media era di 33,2, con un minimo di 21 e un massimo di 52.
[7] Quando le identificazioni non riuscite non vengono incluse nel calcolo, la percentuale di identificazioni corrette è del 100%.

[8] L'amplicone 9 presenta una corsa per omopolimeri di 14 A in base alla sequenza di riferimento del genoma umano versione 19. Tuttavia, le informazioni del riferimento ben caratterizzato per 7 dei 13 campioni presentano 13 A in questa corsa per omopolimeri. In questi 7 campioni, questa delezione di una coppia di basi è chiamata un falso negativo e viene chiamata come falso negativo riproducibile in tutte e nove le corse.

[9] L'amplicone 95 presenta una corsa per omopolimeri di 14 A in base alla sequenza di riferimento del genoma umano versione 19. Tuttavia, le sequenze del riferimento composito ben caratterizzato per 13 dei 13 campioni presentano 15 A in questa corsa per omopolimeri. In questi 13 campioni, questa inserzione di una coppia di basi non è identificata ed è riproducibile al 100% (ad es., è un falso negativo).

I risultati dello Studio 1 sulla riproducibilità per ciascun campione sono mostrati combinati in una colonna da tutte le nove corse. I risultati visualizzati sono solo per le varianti di singolo nucleotide e per i risultati di inserzioni/delezioni rispetto alla sequenza del database di riferimento per tre corse su tre strumenti. Questa analisi dimostra che i risultati per le varianti erano riproducibili, per questi campioni, su nove corse.

Tabella 6 Riepilogo dei risultati di riproducibilità della piattaforma MiSeqDx per 13 campioni ben caratterizzati

N. DNA	ID campione di DNA	N. di corse per campione	N. di SNV	Varianti di singolo nucleotide (SNV)			N. di Indel	Inserzioni/delezioni (Indel)		
				N. di identificazioni corrette	N. di falso positivi[1]	N. di falso negativi[2]		N. di identificazioni corrette	N. di falso positivi[1]	N. di falso negativi[2]
1	NA12877[3]	18	16	16	0	0	3	1	0	2
2	NA12878[3]	18	17	17	0	0	2	0	0	2
3	NA12879	9	18	18	0	0	2	1	0	1
4	NA12880	9	17	17	0	0	3	1	0	2
5	NA12881	9	19	19	0	0	3	1	0	2
6	NA12882	9	15	15	0	0	1	0	0	1
7	NA12883	9	22	22	0	0	2	1	0	1
8	NA12884	9	19	19	0	0	2	1	0	1
9	NA12885	9	17	17	0	0	2	0	0	2
10	NA12886	9	19	19	0	0	3	1	0	2
11	NA12887	9	18	18	0	0	1	0	0	1
12	NA12888	9	22	22	0	0	2	1	0	1
13	NA12893	9	17	17	0	0	3	1	0	2

[1] Falso positivo = Variante identificata dalla corsa di sequenziamento MiSeqDx ma non presente nel database di riferimento.
[2] Falso negativo = Variante presente nel database di riferimento ma non identificata nella corsa di sequenziamento MiSeqDx.
[3] I campioni NA12877 e NA12878 sono stati analizzati in duplicati. I campioni replicati hanno generato risultati identici.

Studio 2

Uno studio di riproducibilità da centro a centro eseguito con un saggio rappresentativo, il saggio MiSeqDx Cystic Fibrosis 139 Variant Illumina, includeva un sottogruppo di variazioni genetiche di *CFTR* significative dal punto di vista clinico analizzate con il software MiSeq Reporter usando il flusso di lavoro per il sequenziamento del DNA target sulla piattaforma MiSeqDx. Lo studio in cieco usava 3 centri di sperimentazione e 2 operatori in ciascun laboratorio. Due pannelli ben caratterizzati di 46 campioni ciascuno sono stati testati da ciascun operatore in ciascun centro per un totale di 810 identificazioni per centro. I panelli erano costituiti da una miscela di DNA genomico proveniente da linee cellulari con varianti note del gene *CFTR*, oltre che da sangue deleucocitato con aggiunta di linee cellulari con varianti note del gene *CFTR*. I campioni di sangue servivano per consentire l'incorporazione delle fasi di estrazione necessarie per preparare il gDNA utilizzato come input primario per il flusso di lavoro del saggio. La percentuale dei campioni "pass", vale a dire il numero di campioni che hanno superato la metrica QC al primo tentativo, è stato del 99,88%. Tutti i risultati del test sono basati su test iniziali.

Tabella 7 Riepilogo dei risultati dello studio di riproducibilità eseguito con un saggio rappresentativo MiSeqDx Cystic Fibrosis 139-Variant

Pannello	N. di campioni	Genotipo del campione	Varianti	Identificazioni totali per centro	Identificazioni con concordanza positiva (varianti)			Identificazioni con concordanza negativa (wild type)			N. di identificazioni errate	N. di identificazioni non riuscite	Concordanza positiva (%)	Concordanza negativa (%)	Concordanza complessiva (%)
					Centro 1	Centro 2	Centro 3	Centro 1	Centro 2	Centro 3					
A	1	S549N (HET)		810	6	6	6	804	804	804	0	0	100	100	100
A	2	1812-1 G>A (HET)		810	6	6	6	804	804	804	0	0	100	100	100
A	3	Q493X/F508del (HET)		810	12	12	12	798	798	798	0	0	100	100	100
A	4[1]	F508del/2184delA (HET)		810	12	12	12	797	798	798	0	1[1]	100	100	100
A	5[2]	Y122X/R1158X (HET)		810	12	10	12	798	665	798	0	135[2]	94,44	94,44	94,44
A	6	F508del/2183AA>G (HET)		810	12	12	12	798	798	798	0	0	100	100	100
A	7	R75X (HET)		810	6	6	6	804	804	804	0	0	100	100	100
A	8	I507del/F508del (HET)		810	12	12	12	798	798	798	0	0	100	100	100

Pannello	N. di campioni	Genotipo del campione	Varianti	Identificazioni totali per centro	Identificazioni con concordanza positiva (varianti)			Identificazioni con concordanza negativa (wild type)			N. di identificazioni errate	N. di identificazioni non riuscite	Concordanza positiva (%)	Concordanza negativa (%)	Concordanza complessiva (%)
					Centro 1	Centro 2	Centro 3	Centro 1	Centro 2	Centro 3					
A	9[3]	F508del/W1282X (HET)		810	12	11	12	798	797	798	2[3]	0	97,22	99,96	99,92
A	10[3]	F508del/3272-26A>G (HET)		810	12	11	12	798	797	798	2[3]	0	97,22	99,96	99,92
A	11	F508del/3849+10kbC>T (HET)		810	12	12	12	798	798	798	0	0	100	100	100
A	12	621+1G>T/3120+1G>A (HET)		810	12	12	12	798	798	798	0	0	100	100	100
A	13	E60X/F508del (HET)		810	12	12	12	798	798	798	0	0	100	100	100
A	14	M1101K (HET)		810	6	6	6	804	804	804	0	0	100	100	100
A	15	M1101K (HOM)		810	6	6	6	804	804	804	0	0	100	100	100
A	16	F508del (HOM)	I506V, I507V, F508C non presente	828	6	6	6	822	822	822	0	0	100	100	100
A	17	F508del/3659delC (HET)		810	12	12	12	798	798	798	0	0	100	100	100
A	18	R117H/F508del (HET)	(TG)10 (T)9/ (TG)12 (T)5	816	18	18	18	798	798	798	0	0	100	100	100

Pannello	N. di campioni	Genotipo del campione	Varianti	Identificazioni totali per centro	Identificazioni con concordanza positiva (varianti)			Identificazioni con concordanza negativa (wild type)			N. di identificazioni errate	N. di identificazioni non riuscite	Concordanza positiva (%)	Concordanza negativa (%)	Concordanza complessiva (%)
					Centro 1	Centro 2	Centro 3	Centro 1	Centro 2	Centro 3					
A	19	621+1G>T/711+1G>T (HET)		810	12	12	12	798	798	798	0	0	100	100	100
A	20	G85E/621+1G>T (HET)		810	12	12	12	798	798	798	0	0	100	100	100
A	21	A455E/F508del (HET)		810	12	12	12	798	798	798	0	0	100	100	100
A	22	F508del/R560T (HET)		810	12	12	12	798	798	798	0	0	100	100	100
A	23	F508del/Y1092X (C>A) (HET)		810	12	12	12	798	798	798	0	0	100	100	100
A	24	N1303K (HET)		810	6	6	6	804	804	804	0	0	100	100	100
A	25	G542X (HOM)		810	6	6	6	804	804	804	0	0	100	100	100
A	26	G542X (HET)		810	6	6	6	804	804	804	0	0	100	100	100
A	27	G551D/R553X (HET)		810	12	12	12	798	798	798	0	0	100	100	100
A	28	3849+10kbC>T (HOM)		810	6	6	6	804	804	804	0	0	100	100	100
A	29	WT		810	0	0	0	810	810	810	0	0	N.D.	100	100
A	30	F508del (HET)		810	6	6	6	804	804	804	0	0	100	100	100
A	31	1717-1G>A (HET)		810	6	6	6	804	804	804	0	0	100	100	100
A	32	R1162X (HET)		810	6	6	6	804	804	804	0	0	100	100	100
A	33	R347P/G551D (HET)		810	12	12	12	798	798	798	0	0	100	100	100

Pannello	N. di campioni	Genotipo del campione	Varianti	Identificazioni totali per centro	Identificazioni con concordanza positiva (varianti)			Identificazioni con concordanza negativa (wild type)			N. di identificazioni errate	N. di identificazioni non riuscite	Concordanza positiva (%)	Concordanza negativa (%)	Concordanza complessiva (%)
					Centro 1	Centro 2	Centro 3	Centro 1	Centro 2	Centro 3					
A	34	R334W (HET)		810	6	6	6	804	804	804	0	0	100	100	100
A	35	WT		810	0	0	0	810	810	810	0	0	N.D.	100	100
A	36	G85E (HET)		810	6	6	6	804	804	804	0	0	100	100	100
A	37	I336K (HET)		810	6	6	6	804	804	804	0	0	100	100	100
A	38	WT		810	0	0	0	810	810	810	0	0	N.D.	100	100
A	39	F508del/3849+10kbC>T (HET)		810	12	12	12	798	798	798	0	0	100	100	100
A	40	621+1G>T/3120+1G>A (HET)		810	12	12	12	798	798	798	0	0	100	100	100
A	41	F508del/3659delC (HET)		810	12	12	12	798	798	798	0	0	100	100	100
A	42	R117H/F508del (HET)	(TG)10 (T)9/ (TG)12 (T)5	816	18	18	18	798	798	798	0	0	100	100	100
A	43	G85E/621+1G>T (HET)		810	12	12	12	798	798	798	0	0	100	100	100
A	44	A455E/F508del (HET)		810	12	12	12	798	798	798	0	0	100	100	100
A	45	N1303K (HET)		810	6	6	6	804	804	804	0	0	100	100	100
A	46	G551D/R553X (HET)		810	12	12	12	798	798	798	0	0	100	100	100

Pannello	N. di campioni	Genotipo del campione	Varianti	Identificazioni totali per centro	Identificazioni con concordanza positiva (varianti)			Identificazioni con concordanza negativa (wild type)			N. di identificazioni errate	N. di identificazioni non riuscite	Concordanza positiva (%)	Concordanza negativa (%)	Concordanza complessiva (%)
					Centro 1	Centro 2	Centro 3	Centro 1	Centro 2	Centro 3					
B	47	2789+5G>A (HOM)		810	6	6	6	804	804	804	0	0	100	100	100
B	48	CFTR dele2, 3/F508del (HET)		810	12	12	12	798	798	798	0	0	100	100	100
B	49	F508del/1898+1G>A (HET)		810	12	12	12	798	798	798	0	0	100	100	100
B	50	WT		810	0	0	0	810	810	810	0	0	N.D.	100	100
B	51	F508del/2143delT (HET)		810	12	12	12	798	798	798	0	0	100	100	100
B	52	3876delA (HET)		810	6	6	6	804	804	804	0	0	100	100	100
B	53	3905insT (HET)		810	6	6	6	804	804	804	0	0	100	100	100
B	54	394delTT (HET)		810	6	6	6	804	804	804	0	0	100	100	100
B	55	F508del (HET)		810	6	6	6	804	804	804	0	0	100	100	100
B	56	WT		810	0	0	0	810	810	810	0	0	N.D.	100	100
B	57	WT		810	0	0	0	810	810	810	0	0	N.D.	100	100
B	58	F508del (HET)		810	6	6	6	804	804	804	0	0	100	100	100
B	59	WT		810	0	0	0	810	810	810	0	0	N.D.	100	100
B	60	L206W (HET)		810	6	6	6	804	804	804	0	0	100	100	100
B	61	WT		810	0	0	0	810	810	810	0	0	N.D.	100	100

Pannello	N. di campioni	Genotipo del campione	Varianti	Identificazioni totali per centro	Identificazioni con concordanza positiva (varianti)			Identificazioni con concordanza negativa (wild type)			N. di identificazioni errate	N. di identificazioni non riuscite	Concordanza positiva (%)	Concordanza negativa (%)	Concordanza complessiva (%)
					Centro 1	Centro 2	Centro 3	Centro 1	Centro 2	Centro 3					
B	62	G330X (HET)		810	6	6	6	804	804	804	0	0	100	100	100
B	63	WT		810	0	0	0	810	810	810	0	0	N.D.	100	100
B	64	R347H (HET)		810	6	6	6	804	804	804	0	0	100	100	100
B	65	1078delT (HET)		810	6	6	6	804	804	804	0	0	100	100	100
B	66	G178R/F508del (HET)		810	12	12	12	798	798	798	0	0	100	100	100
B	67	S549R (c.1647T>G) (HET)		810	6	6	6	804	804	804	0	0	100	100	100
B	68	S549N (HET)		810	6	6	6	804	804	804	0	0	100	100	100
B	69	W846X (HET)		810	6	6	6	804	804	804	0	0	100	100	100
B	70	WT		810	0	0	0	810	810	810	0	0	N.D.	100	100
B	71	E92X/F508del (HET)		810	12	12	12	798	798	798	0	0	100	100	100
B	72[4]	621+1G>T/1154insTC (HET)		810	12	12	12	798	798	797	0	1[4]	100	99,96	99,96
B	73	G542X (HET)		810	6	6	6	804	804	804	0	0	100	100	100
B	74	F508del (HET)		810	6	6	6	804	804	804	0	0	100	100	100
B	75[2]	F508del (HET)		810	6	5	6	804	670	804	0	135[2]	94,44	94,44	94,44
B	76	F508del (HET)		810	6	6	6	804	804	804	0	0	100	100	100

Pannello	N. di campioni	Genotipo del campione	Varianti	Identificazioni totali per centro	Identificazioni con concordanza positiva (varianti)			Identificazioni con concordanza negativa (wild type)			N. di identificazioni errate	N. di identificazioni non riuscite	Concordanza positiva (%)	Concordanza negativa (%)	Concordanza complessiva (%)
					Centro 1	Centro 2	Centro 3	Centro 1	Centro 2	Centro 3					
B	77	621+1G>T/A455E (HET)		810	12	12	12	798	798	798	0	0	100	100	100
B	78	1812-1 G>A (HET)		810	6	6	6	804	804	804	0	0	100	100	100
B	79	WT		810	0	0	0	810	810	810	0	0	N.D.	100	100
B	80	F508del/R553X (HET)		810	12	12	12	798	798	798	0	0	100	100	100
B	81	F508del/G551D (HET)		810	12	12	12	798	798	798	0	0	100	100	100
B	82	R347P/F508del (HET)		810	12	12	12	798	798	798	0	0	100	100	100
B	83	R117H/F508del (HET)	(TG)10 (T)9/ (TG)12 (T)5	816	18	18	18	798	798	798	0	0	100	100	100
B	84	I507del (HET)		810	6	6	6	804	804	804	0	0	100	100	100
B	85	2789+5G>A (HOM)		810	6	6	6	804	804	804	0	0	100	100	100
B	86[4]	CFTR dele2, 3/F508del (HET)		810	12	12	12	798	797	798	0	1[4]	100	99,96	99,96
B	87	F508del/1898+1G>A (HET)		810	12	12	12	798	798	798	0	0	100	100	100
B	88	WT		810	0	0	0	810	810	810	0	0	N.D.	100	100

Pannello	N. di campioni	Genotipo del campione	Varianti	Identificazioni totali per centro	Identificazioni con concordanza positiva (varianti)			Identificazioni con concordanza negativa (wild type)			N. di identificazioni errate	N. di identificazioni non riuscite	Concordanza positiva (%)	Concordanza negativa (%)	Concordanza complessiva (%)
					Centro 1	Centro 2	Centro 3	Centro 1	Centro 2	Centro 3					
B	89	F508del/2143delT (HET)		810	12	12	12	798	798	798	0	0	100	100	100
B	90	3905insT (HET)		810	6	6	6	804	804	804	0	0	100	100	100
B	91	394delTT (HET)		810	6	6	6	804	804	804	0	0	100	100	100
B	92	F508del (HET)		810	6	6	6	804	804	804	0	0	100	100	100
totali				74556	2209			221182			4	273	99,77	99,88	99,88

[1] La posizione del wild type corrispondente alla variante N1303K per un replicato ha prodotto un'identificazione non riuscita a causa della coperture insufficiente.
[2] Un replicato dei campioni 5 e 75 ha registrato una percentuale di identificazione dello 0%. Ulteriore investigazione indica che i campioni potrebbero non essere stati aggiunti alla piastra campioni prima della preparazione della libreria perché i volumi dei campioni rimanenti nelle provette erano coerenti con nessun volume rimosso.
[3] Evidenze empiriche indicano che probabilmente i campioni 9 e 10 sono stati scambiati dall'operatore prima della preparazione della libreria.
[4] La posizione del wild type corrispondente alla variante M1V per un replicato di ciascuno dei due campioni ha prodotto un'identificazione non riuscita a causa della copertura insufficiente.

Estrazione del DNA

Tre diversi metodi di estrazione, estrazione con microsfere magnetiche, precipitazione alcolica e isolamento mediante colonna di gel di silice, sono stati valutati utilizzando sangue intero anticoagulato in K_2EDTA. In questo studio sono stati usati quattordici campioni di sangue unici che rappresentano una gamma di genotipi da un gene rappresentativo. I tre metodi di estrazione del DNA sono stati analizzati indipendentemente da 2 diversi operatori e ciascuno di loro ha eseguito 3 corse per ciascun metodo di estrazione. Ciascuna estrazione è stata eseguita da ciascun operatore in giorni diversi. La concentrazione di DNA e il rapporto A260/A280 dei campioni di gDNA estratto sono stati determinati usando spettrofotometria. La dimensione complessiva del campione per ciascun metodo di estrazione esaminato nello studio è stato pari a 168 (14 campioni x 2 operatori/metodo di estrazione x 3 corse/operatore x 2 replicati/campioni di gDNA estratto).

Metodo di estrazione	Numero di campioni analizzati	Percentuale di identificazione	Accuratezza[1]	Percentuale di campioni "first pass"[2]
Precipitazione alcolica	168	100%	100%	100%
Isolamento su colonna di gel di silice	168	100%	100%	100%
Estrazione con microsfere magnetiche	168	100%	100%	100%

[1]Accuratezza - La percentuale di concordanza con un metodo di analisi di riferimento (sequenziamento bidirezionale Sanger) calcolata per quelle posizioni delle basi che hanno ottenuto un'identificazione delle basi.
[2]Percentuale di campioni "first pass" - Il numero di campioni che corrispondono alla percentuale di identificazione specificata la prima volta che sono elaborati (senza la necessità di ripetere la corsa o elaborazione ulteriore) sottoforma di una percentuale del numero totale dei campioni analizzati durante un singolo esperimento di sequenziamento MiSeqDx.

Input di DNA

La gamma di input di DNA per la piattaforma MiSeqDx è stata valutata eseguendo uno studio di diluizione in serie usando 14 campioni di DNA rappresentativi che contenevano 16 singole varianti del gene uniche. Ciascun campione è stato analizzato in duplicati a 9 livelli di input di DNA che andavano da 1250 ng a 1 ng (1250 ng, 500 ng, 250 ng, 100 ng, 50 ng, 25 ng, 10 ng, 5 ng e 1 ng). Per la determinazione dell'accuratezza, i genotipi dei campioni sono stati confrontati con i dati ottenuti dal sequenziamento bidirezionale Sanger. 1250 ng e 25 ng sono stati identificati come il legame superiore e inferiore per l'input di DNA rispettivamente in quanto hanno ottenuto una percentuale di campioni "first pass" del ≥95% senza identificazioni non corrette (100% di accuratezza e percentuale di identificazione).

Gli input di DNA di 1250 ng, 250 ng e 100 ng sono stati ulteriormente analizzati con 4 campioni di DNA rappresentativi e 20 replicati per ciascun livello di input di DNA per ciascun campione (n=4*20=80 campioni), mentre il legame inferiore di 25 ng è stato analizzato con 14 campioni, 20 replicati per ciascun campione (n=14*20=280 campioni). L'accuratezza e la percentuale di campioni "first pass" sono risultati pari al 100% a tutti i livelli di input di DNA e percentuale di identificazione dei campioni di >99%.

Sostanze interferenti

Per valutare l'impatto delle sostanze interferenti sulla piattaforma MiSeqDx, un saggio rappresentativo progettato per interrogare un gene singolo che copre 11.529 basi è stato valutato in presenza o in assenza di potenziali sostanze interferenti. Nello studio sono stati usati otto campioni di sangue intero che rappresentavano otto genotipi unici. Quattro sostanze interferenti endogene (bilirubina, colesterolo, emoglobina e trigliceride) sono state testate aggiungendole ai campioni di sangue prima dell'estrazione del DNA. Per valutare l'interferenza risultante dalla

raccolta del sangue (prelievo breve), EDTA è stato aggiunto ai campioni di sangue in due concentrazioni. I limiti di concentrazione di ciascuna sostanza sono riportati nella tabella seguente. Inoltre, per valutare l'interferenza risultante dalla preparazione dei campioni, è stato aggiunto 15% di tampone di lavaggio a 8 DNA genomici purificati. È stata raggiunta una percentuale di identificazione del 100% per tutti i campioni analizzati oltre a una riproducibilità del 100% nell'identificazione dei genotipi tra i campioni in presenza o in assenza delle sostanze interferenti.

Sostanza del test	Numero totale di replicati	Concentrazione analizzata nel sangue (limite superiore)	Concentrazione analizzata nel sangue (limite inferiore)	Percentuale di identificazione
Bilirubina	16	684 μmol/l	137 μmol/l	100%
Colesterolo	16	13 mmol/l	2,6 mmol/l	100%
Emoglobina	16	2 g/l	0,4 g/l	100%
Trigliceride	16	37 mmol/l	7,4 mmol/l	100%
EDTA	16	7 mg/ml	2,8 mg/ml	100%

Indicizzazione del campione

I primer indice dei campioni sono utilizzati nel kit per assegnare un codice a barre univoco a ciascun campione di DNA, consentendo di raggruppare in un pool più campioni in una singola corsa di sequenziamento.

Sono stati analizzati complessivamente 96 indici di campioni mediante un saggio rappresentativo progettato per interrogare un gene singolo che copre 11.529 basi con 8 campioni di DNA unico al fine di verificare la capacità del saggio di identificare i genotipi in modo coerente per un dato campione fra diverse combinazioni di primer di indicizzazione. Ciascun campione è stato analizzato con 12 diverse combinazioni di primer di indicizzazione. Quarantotto (48) combinazione di indici sono state analizzate in una corsa di sequenziamento. I risultati dei campioni sono stati confrontati con i dati ottenuti dal sequenziamento bidirezionale Sanger per tutte le posizioni/varianti. Riproducibilità e accuratezza sono risultate del 100% per tutte le combinazioni di primer indice/campione.

Brevetti e marchi registrati

Informazioni di contatto

Illumina
San Diego, California 92122 U.S.A.
+1.800.809.ILMN (4566)
+1.858.202.4566 (fuori dal Nord America)
techsupport@illumina.com
www.illumina.com

Emergo Europe
Molenstraat 15
2513 BH L'Aia
Paesi Bassi

Etichettatura del prodotto

Fare riferimento alla chiave dei simboli spediti con ciascun kit per un riferimento completo ai simboli che possono apparire sulla confezione e sull'etichettatura del prodotto.